A Data Science Solution for Measurement and Verification 2.0 in Industrial Buildings

Colm Vincent Gallagher
BE (Hons)

Thesis submitted for the degree of Doctor of Philosophy

NATIONAL UNIVERSITY OF IRELAND, CORK

School of Engineering

March 2019

Head of School: Prof. Liam Marnane

Supervisor: Dr. Dominic O'Sullivan

Research supported by the NTR Foundation and Science Foundation Ireland MaREI Centre

Contents

List of Figures

List of Tables

List of Acronyms

R^2 coefficient of determination.

$R^2_{adjusted}$ adjusted coefficient of determination.

AFDD automated fault detection and diagnosis.

AHU air handling unit.

AI artificial intelligence.

AMI advanced metering infrastructure.

ANN artificial neural network.

API application programming interface.

ASHRAE American Society of Heating, Refrigerating and Air-Conditioning Engineers.

AWS Amazon Web Services.

.

BMS building management system.

BOD Basis of Design.

CART classification and regression tree.

CDD cooling degree days.

CFD computational fluid dynamics.

CHP combined heat and power.

CV coefficient of variation.

CV(RMSE) cofficient of variation of the root mean squared error.

DOE Department of Energy.

EC European Commission.

ECM energy conservation measure.

ECR Elastic Container Registry.

EED Energy Efficiency Directive.

EMS energy management system.

EnMS energy management system.

EPC energy performance contracting.

ESCO energy services company.

EU European Union.

EVO Efficiency Valuation Organization.

FDD fault detection and diagnosis.

GA genetic algorithm.

GDP Gross Domestic Product.

GP Gaussian process.

GUI graphical user interface.

HDD heating degree days.

HVAC heating, ventilation and air-conditioning.

IERG Intelligent Efficiency Research Group.

IPMVP International Performance Measurement and Verification Protocol.

IQR inter quartile range.

ISO International Organization for Standardization.

k-NN k-nearest neighbours.

kW kilowatts.

M&T monitoring and targetting.

M&V measurement and verification.

M&V 2.0 measurement and verification 2.0.

med(absRTE) median of the absolute relative error.

NMBE normalised mean bias error.

OAT outside air temperature.

OLS ordinary least-squares regression.

PCA principle component analysis.

PDCA Plan-Do-Check-Act.

PDD performance deviation detection.

PMML Predictive Model Markup Language.

PRIM patient rule induction method.

REF2016 Reference Scenario 2016.

REST Representational State Transfer.

RMSE root mean squared error.

RO research objective.

RSS residual sum of squares.

S3 Simple Storage Service.

SE standard error.

SEP Superior Energy Performance.

SMEs small and medium-sized enterprises.

SVM support vector machine.

TFEC total final energy consumption.

TPES total primary energy supply.

U.S. United States.

URL uniform resource locator.

VIF variance inflation factor.

VPC Virtual Private Cloud.

VSD variable speed drive.

I, Colm Vincent Gallagher, certify that this thesis is my own work and has not been submitted for another degree at University College Cork or elsewhere. All external references and sources are clearly acknowledged and identified within the contents. I have read and understood the regulations of University College Cork concerning plagiarism.

Colm Vincent Gallagher

"Not everything that counts can be counted, and not everything that can be counted counts"

- William Bruce Cameron

Executive Summary

As the mix of fuels satisfying the ever growing worldwide energy demand changes, end-use energy systems are also undergoing a transition. This presents a number of challenges, which the advent of digitalisation and Industry 4.0 practices can offer solutions to. The December 2015 Paris Agreement and national energy policies demonstrate a global commitment to addressing the negative impact humans have on the environment. On the demand-side, the threat of rising energy prices, the cost benefits arising from lower consumption and public financing are driving investment in energy efficiency improvements. However, there exists a number of barriers that are preventing investments in cost-effective measures. This has resulted in an energy efficiency gap being established. Risk, uncertainty and hidden costs all contribute to this difference between the optimal and actual levels of investment in energy efficiency. With global efforts focused on closing this gap and the increased reliance on energy efficiency as a resource, the accounting system used for measuring and verifying energy savings has been brought into question. Measurement and verification (M&V) is this accounting system and it exists as a sub-sector of the energy industry. In developing solutions to close the energy efficiency gap, M&V plays a central role in overcoming the barriers to investment that exist.

The advent of advanced metering infrastructure has led to vast quantities of energy data becoming available. Despite this, the typical methods employed for the performance verification of energy efficiency improvements have not progressed, as they continue to rely on expert knowledge and simplistic statistical modelling techniques. This leads to uncertainty in the quantity of savings arrived at, with this uncertainty acting as a barrier to investment in energy efficiency. In response to this, the industry is evolving towards more advanced and automated methods known as M&V 2.0. This however presents the challenge of keeping the resources required to perform M&V at a minimum level, while also improving the accuracy, reliability and trust in the process.

The research presented in this thesis can be largely classified into two prominent tasks. These are the development of a machine learning-based methodology for the construction of accurate baseline energy models and the establishment of a framework and final solution for M&V 2.0. It will be shown through theoretical and practical work that:

- Machine learning techniques reduce the uncertainty introduced into the

performance verification process by the baseline energy regression model. Additionally, the utilisation of a broader scope of analysis with respect to traditional methods is advantageous in further improving model accuracy (Gallagher, Bruton, Leahy & O'Sullivan 2018).

- Novel, computationally efficient data processing methods, including cleaning and feature selection, can be tailored for the industrial buildings sector to minimise the resources required to carry out performance verification (Gallagher, Leahy, O'Donovan, Bruton & O'Sullivan 2018).
- The void in knowledge resulting from the established M&V protocols can be populated by a prescriptive methodology that utilises machine learning techniques to accurately and reliably quantify energy savings; thus, empowering performance verification practitioners in the use of advanced analytics on granular data sets and removing the need for expert knowledge (Gallagher, Leahy, O'Donovan, Bruton & O'Sullivan 2018).
- The industrial sector requires a specific framework for the application of M&V 2.0 practices. An M&V 2.0 framework is developed to offer a solution to the challenge of persisting energy savings. A performance deviation detection system enables integration with ongoing monitoring and targeting practices (Gallagher, O'Donovan, Leahy, Bruton & O'Sullivan 2018).
- M&V 2.0 does not have to increase the resources required to carry out performance verification. A novel, cloud computing-based solution, IntelliMaV, is capable of quantifying energy savings in near real-time with minimal uncertainty at high confidence levels (Gallagher et al. 2019).

This thesis addresses some of the key challenges facing the performance verification industry including utilising the large quantities of energy data available in industrial facilities and evolving practices to a level of maturity that will enable it to support M&V 2.0. The implications of such challenges are shown to be significant beyond the individual project level, with the effectiveness of European energy policy dependent on accurate and reliable M&V. The methodology, framework and IntelliMaV application developed all address these challenges, while aiding the transition to M&V 2.0 practices. Despite these advancements, this is not the final solution for the industry. A collective effort must be made to continue to modernise performance verification practices to ensure M&V remains a valued and beneficial practice in energy management into the future.

Acknowledgements

I am indebted to a number of people who have offered advice, guidance and support throughout my postgraduate studies. I would like to give a special mention to those that without whom, I could not have completed this thesis.

Firstly, I would like to extend my sincere gratitude to my supervisor, Dr. Dominic O'Sullivan, for affording me the opportunity to undertake this research and for his mentorship over the past number of years. His open door policy and constant willingness to offer advice and help in developing solutions to the challenges faced along the way meant I could not have asked for any more from a supervisor.

I would also like to thank Prof. Brian Ó Gallachóir and Dr. Fionn Rogan for their work on the 'Opportunities for Ireland in a Low Carbon Economy - Our 2050' project. The funding received from the NTR Foundation and Science Foundation Ireland allowed us to work together on a compelling and well received body of research. With Fionn leading the team admirably, it was thoroughly enjoyable to work with talented colleagues towards shared objectives.

I found the support offered by MaREI to be invaluable throughout my time spent researching in UCC. Angela, Gillian, Peter and Tara, thank you so much for not just helping me with funding, but also being a friendly port of call when I ventured down to Ringaskiddy. Thanks also to Sheila, Richie and Julie in the School of Engineering.

My colleagues in the Intelligent Efficiency Research Group (IERG) have been a fantastic source of knowledge and were always happy to give input on my research. Ken, Kev, Peter, Brendan, Seán (x2) and our new members, I greatly appreciate the time you have spent aiding me on my research path and I am privileged to have collaborated with you all.

I would like to pay particular thanks to everyone in CE203, past and present, for creating such a good atmosphere to work in. There is a long list of you but I will give it my best shot: Brendan, Ciarán, Kev, Kieran, Paul, Seán Blake, Ken, Deirdre, Paddy, Annie, Lara, David, Di, Seán Hayes, John, Romain and the many others who have come and gone during my time. I have really enjoyed my years spent in UCC with you all, be it over coffee or pints.

On a personal note, I cannot but thank my parents, Norah and Vince, and my sister, Niamh, for everything they have done for me. I have been encouraged and instilled with confidence each and every step of the way. Mom and Dad, I won't say your job is done, but the workload should certainly be a lot lighter.

Chapter 1

Introduction

1.1 Background and Motivation

1.1.1 Growing Global Energy Demand

For the past 45 years, the worldwide total primary energy supply (TPES) has grown consistently with an increase of almost 150% from 1971 to 2016 (International Energy Agency 2018). The mix of fuels satisfying this demand has evolved over this period in response to economic, political and environmental conditions. This has resulted in global energy systems undergoing significant changes as countries seek to transition towards low carbon economies and reduce the negative environmental impacts caused by modern society. This transition presents extensive multi-disciplinary challenges that need to be overcome to enable full realisation of a low carbon global energy system. These challenges include the integration of disruptive renewable energy technologies into grids that historically lack dynamism, sustainable means of satisfying increasing energy demands and reducing waste across all aspects of our energy systems.

The advent of globalised energy systems offers a unique opportunity to collaboratively develop solutions to the problems faced in reducing the negative impact humans have on the environment. The December 2015 Paris Agreement demonstrates a commitment to this with 180 countries to date ratifying a target to keep global temperature rise this century below 2°C above pre-industrial levels and to pursue efforts to limit the temperature increase even further to 1.5°C (United Nations Framework Convention on Climate Change 2015). This agreement reinstates a firm global commitment to minimising the changes to the Earth's climate system. Extensive action is required on an international, national and regional scale to ensure the successful attainment of this target. Strides are being made through these actions at present, although continued, targeted and properly evaluated policy is required in the future. As TPES has increased by almost 150% since 1971, total final energy consumption (TFEC) has grown by 125% in the same period (International Energy Agency 2018). However, improvements are being achieved with global energy intensity, defined as the TFEC per unit of economic output, decreasing over the past 6 years (Figure 1.1). This decreasing trend, which began two decades ago, is evidence of energy being consumed in a more efficient manner worldwide and has a critical role in curtailing the consequences of an ever increasing energy demand.

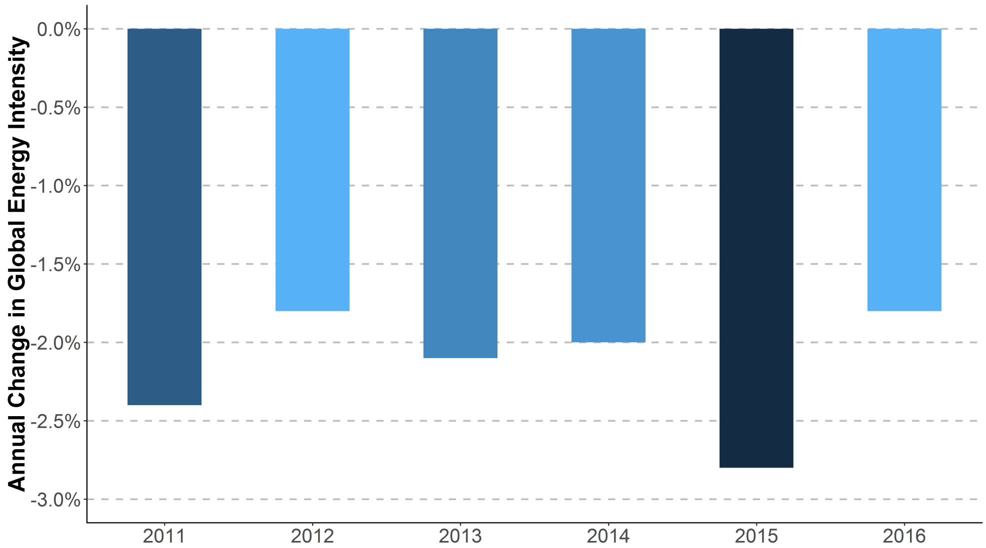


Figure 1.1: Annual changes in global energy intensity (energy per unit of GDP) (Data source: International Energy Agency (2017))

Despite the de-carbonisation of energy supply through the use of renewable energy sources, it is critical that energy is consumed with maximum efficiency on the demand-side to reduce waste, cost and energy intensity. Figure 1.2 shows the sectoral breakdown of TFEC in 2016. As the sectoral breakdown of TFEC has transitioned since 1971 due to the evolution of worldwide energy systems, energy use in the industrial sector has consistently dominated TFEC for this period and accounted for 37% of global energy consumption in 2016. Ambitious energy policies are required to slow the rate at which global TFEC is increasing and alter the trajectory of projections for the betterment of society. The successful implementation of such policies would require significant progress to be made in the industrial sector.

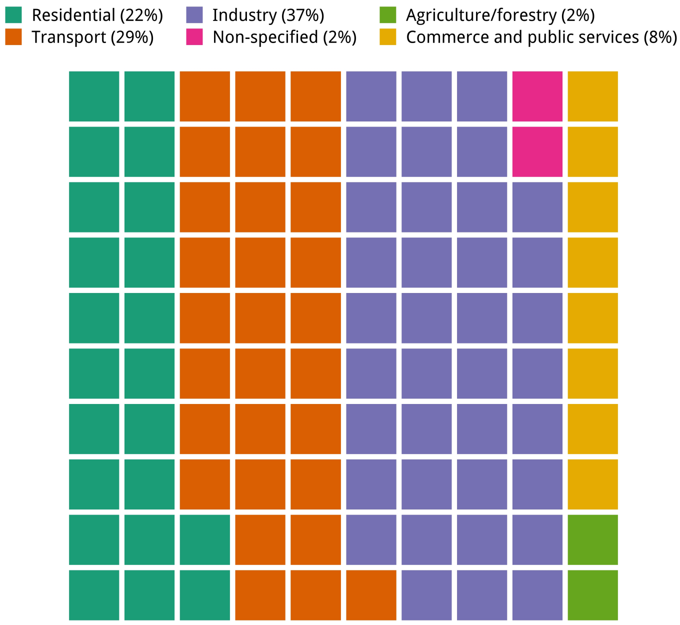


Figure 1.2: Global total final energy consumption by sector in 2016 (Data source: International Energy Agency (2018))

1.1.2 The Energy Efficiency Resource

Energy efficiency is essential in ensuring a safe, reliable, affordable and sustainable energy system for the future. It is a resource possessed by every country in abundance and offers a least-cost means of addressing energy security, environmental and economic challenges (American Council for an Energy Efficieny Economy 2018). There remains a vast portion of the energy efficiency opportunity that has yet to be realised, often referred to as the energy efficiency gap. This is the discrepancy between the optimal and actual implementation of cost-effective energy conservation measures (ECMs) (Jaffe & Stavins 1994). The reasons for this gap and solutions to address it are discussed in detail in Chapter 2, but prior to that, it is important to note the value in such opportunities.

Improvements in energy efficiency are the biggest contributor to reductions in energy consumption and associated emissions, with it accounting for more than double the impact caused by the shift in economic activity towards less energy-intensive sectors. Crucially, in 2016, the world would have used 12% more energy had the efficiency improvements implemented since 2000 not been realised (International Energy Agency 2017). This is equivalent to adding another European Union (EU) to the global energy market. With energy efficiency being the main driver in reducing end-use demand, it is a valuable resource in the transition towards a low-carbon future. Figure 1.3 illustrates the economic value of improved energy intensity (in Billion US Dollars) in 2016 alone, with energy efficiency being a critical tool in generating this value. These figures are based on difference between actual Gross Domestic Product (GDP) and its notional level that would have been generated had energy intensity stayed the same since the previous year.

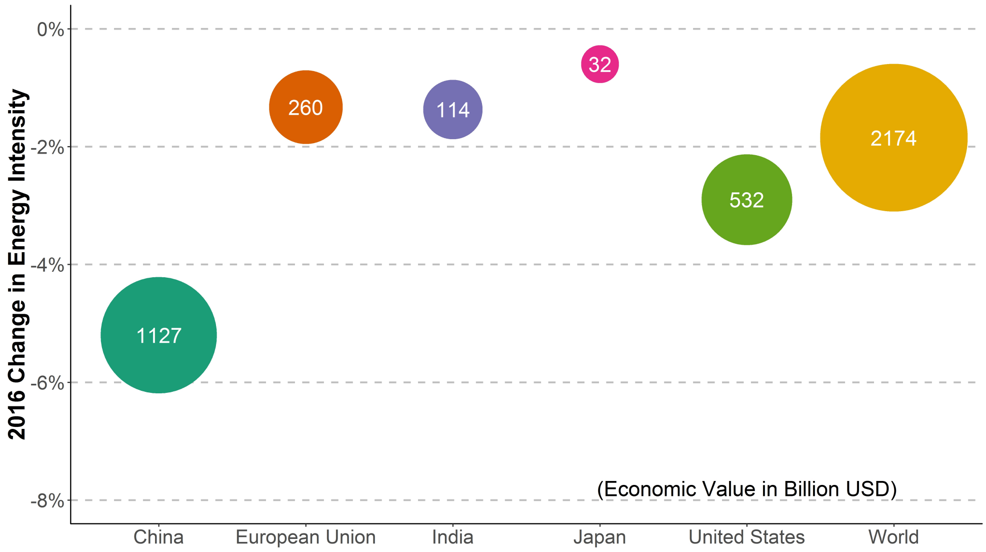


Figure 1.3: Economic value of improved energy intensity (Data source: International Energy Agency (2017))

There is large scope for continuing these improvements in energy intensity through the closing of the energy efficiency gap. The EU Reference Scenario 2016 (REF2016) provides projections that show a possible future energy system for member states (Capros et al. 2016). It does so by taking into account global and EU market trends and the energy and climate policies already adopted by the EU and its Member States. A set of assumptions on variables such as population growth, macroeconomic and oil price developments, technology improvements, and policies are also applied. Importantly, REF2016 projects significant energy efficiency improvements to be a main resource in reducing the EU's import dependency as fossil fuel production decreases. The scenario expects energy policy to be the primary driver for efficiency improvements up to 2020, with market trends taking control following this. In a pivotal move towards addressing the sectoral shares in TFEC (Figure 1.2), demand in industry is projected to decrease by 5% in the EU. This is mostly due to improved energy efficiency in non-energy intensive industries. Figure 1.4 illustrates the further decoupling of value added and final energy consumption in the industrial sector. In the long term, the projected improvements are a result of embedding energy efficient technologies in industry in place of old equipment and a shift towards higher value added and less-energy intensive production processes.

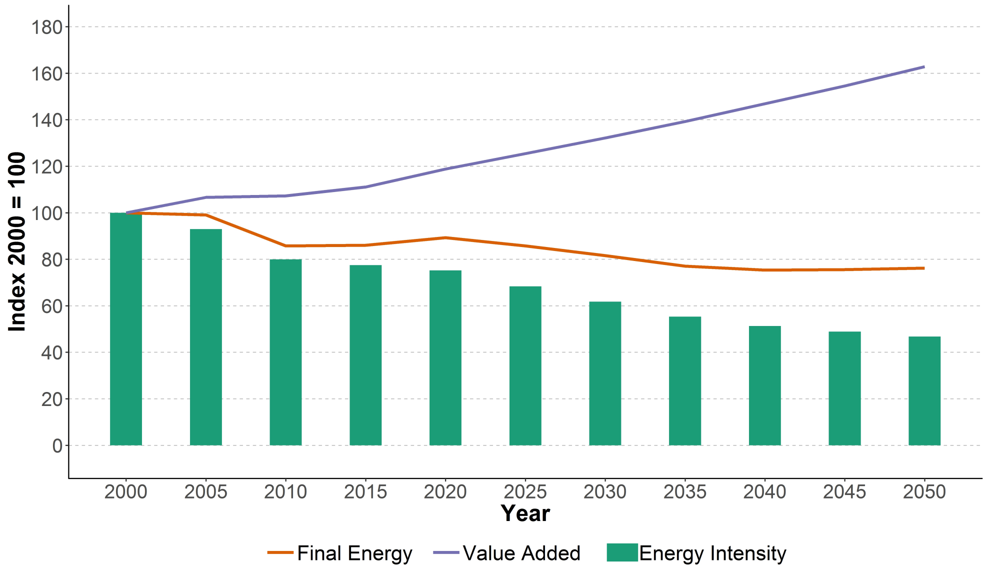


Figure 1.4: Industrial energy demand versus activity (value added) (Data source: Capros et al. (2016))

Projections towards the desired low-carbon economies of the future consistently identify energy efficiency as a key actor in the transition. However, there are technical challenges facing energy efficiency that must be overcome to fully realise the opportunity. As projected by REF2016, energy policy plays an initial key role in ensuring solutions to these challenges are developed and the benefits resulting from implementation are realised.

1.1.3 European Energy Efficiency Policy

In 2007, the European Council ratified a progressive set of energy targets for the coming years (Council of the European Union 2007). By 2020, the EU aims to reduce its greenhouse gas emissions by at least 20%, increase the share of renewable energy to at least 20% of consumption and achieve energy savings of 20% or more. The energy savings target is to be realised by incentivising efficiency measures through policy initially. Progress towards the 20% savings target is evaluated by comparing actual consumption to the projected use of energy in 2020.

However, forecasts in 2010 signalled that the EU would not meet its 2020 target for energy efficiency. In response to this, the 2012 Energy Efficiency Directive (EED) establishes a set of binding measures pursuant to the 2020 energy efficiency target. Each Member State has varying available resources and their own unique energy markets, which results in each requiring a distinctive course of action when it comes to meeting their obligations under the EED. The Directive requires Member States to set national energy efficiency targets for 2020. These targets were evaluated in terms of their contribution to meeting the EU's overall target and the extent to which individual countries meet the common goal.

A key requirement of Member states is the implementation of an energy efficiency obligation scheme, or alternative policy measures, to ensure energy savings are realised at an end-use level. This component (Article 7 of the EED) accounts for half of the energy savings the Directive is designed to achieve (European Parliament and Council 2012). Member states are obligated to achieve new savings every year of at least 1.5% of the annual energy sales to final customers of all energy distributors or all retail energy sales companies by volume averaged over 2010, 2011 and 2012. This quantity of new energy savings are required every year from 1 January 2014 through to 31 December 2020. There are a number of permutations in existence that allow Member States to implement this using differing approaches, however, savings must not be reduced by more than 25% of the overall target. Therefore, member states must achieve a minimum of 75% of their 1.5% annual target through the implementation of an energy efficiency obligation scheme or appropriate alternative policy measures.

The foundations of the measures implemented pursuant to the terms of the EED are based on energy savings persisting beyond their fist year of realisation. Therefore, the benefits of ECMs implemented in 2014 must still be realised in 2020. Table 1.1 presents an example of such process using a sample new energy savings target of 1.5%. The successful implementation of such a system requires energy savings to be accurately measured and verified over the duration of the period for which they are counted towards EU energy efficiency targets. This energy savings accounting system requires the use of measurement and verification (M&V), with Paragraph 6 of Article 7 ensuring this is conducted by independent parties for a statistically significant proportion and representative sample of the ECMs implemented. M&V is the term given to the process of quantifying energy savings resulting from the implementation of an ECM. It exists as a performance verification sub-sector of the energy industry. Annex V of the EED sets out a range of methods and principles which need to be followed in the calculation of energy savings, all of which are reliant on M&V.

In June 2018, the European Commission, Parliament and Council reached a political agreement that includes a binding energy efficiency target for the EU for

Table 1.1: Example of cumulative energy savings approach required by the EED

Year	New Energy Savings	Cumulative Energy Savings
1	1.5%	1.5%
2	1.5%	3%
3	1.5%	4.5%
4	1.5%	6%
5	1.5%	7.5%
6	1.5%	9%
7	1.5%	10.5 %

2030 of 32.5%. Additionally, a clause is included that allows for an upwards revision of this target by 2023 (European Commission 2018). A new Directive is being prepared to establish obligations under which Member States must comply to ensure the EU achieves this ambitious target by 2030. It is expected that this Directive will further increase the dependency on M&V as a vital tool in the transition to low-carbon energy systems.

It is evident that a multi-faceted approach is required to achieve the energy efficiency targets set by the European Commission (EC) for both 2020 and 2030. The reliable and accurate quantification of energy savings at an end-use level is a critical step in the delivery of over half the savings the EED is designed to deliver. Therefore, a detailed review of the challenges facing M&V must be undertaken to identify the barriers to successful implementation of the measures that Member States are legally obligated to pursue. Inaccurate M&V has significant potential to proclaim untrue progress in delivering energy efficiency savings, thus negatively impacting on the effective delivery of energy policy.

1.1.4 Challenges Facing Measurement and Verification

There are three periods of interest in any M&V project: the baseline, implementation and reporting periods. Figure 1.5 provides an illustration of the overall process that is typically applied. Although they always occur sequentially, the length of each period will vary depending on individual project parameters. The baseline period occurs prior to the implementation of an ECM, with the reporting period taking place following the implementation period. A crucial step in M&V is the estimation of the adjusted baseline in the reporting period. This is found by normalising the reporting period energy consumption to baseline period conditions. Typically, engineering or statistical methods are applied to construct a baseline energy model capable of performing this normalisation. Consequently, M&V is not an exact science and maintaining accuracy throughout the process is critical to its success. A mathematical description of the M&V problem provides an insight into the concepts, techniques and methodologies in M&V (Xia & Zhang 2013).

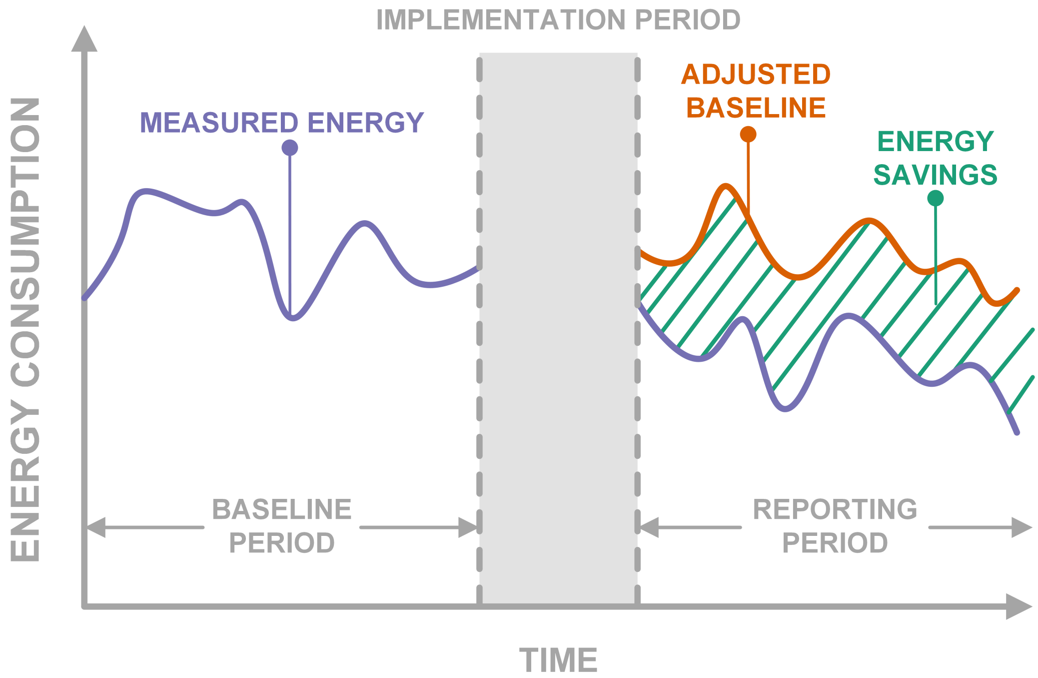


Figure 1.5: Graphical illustration of measurement and verification process

The three principle sources of uncertainty in the process are sampling, measuring and modelling. One focal point of this thesis is addressing the uncertainty introduced in the development of the baseline energy model used to estimate the adjusted baseline. With the increasing availability of large quantities of energy data from advanced metering infrastructure (AMI), there exists an opportunity to vastly improve the traditional methods used to construct these baseline energy models. The significance of ensuring accuracy in this task and thus, minimising the uncertainty in the energy savings quantified, is evident given the impact individual ECMs have on macro-level energy efficiency targets. The realisation of these targets is reliant on the successful implementation of ECMs on an individual project level. For each project, M&V is used to deliver a quantification of savings with an associated level of uncertainty for a given confidence interval. The sum of these independent quantities of savings is used to evaluate progress against higher-level targets. Consequently, there will always be uncertainty associated with the National and European quantities of energy savings. Accomplishing the measures set out in the EED necessitates the minimisation of uncertainty in each and every quantification of energy savings. Failure to do so will result in inaccurate performance evaluation of energy efficiency improvement measures, thus preventing effective action to be taken on a European, National and regional level.

In recent years, measurement and verification 2.0 (M&V 2.0) represents an area of significant growth and it is being used to further develop the commonly used practices. M&V 2.0 differs from traditional M&V as it uses large data sets and automated advanced analytics to streamline and scale the process (Granderson et al. 2017). The automated analytics can provide ongoing savings estimates in close to real-time. This enables M&V to progress from a static, retrospective process to a more dynamic state in which savings can be maximised. This is achievable by identifying deviations in performance and subsequently, rectifying errors in ECM implementation and operation. The added complexity of the baseline energy model has been driven by the increased availability of granular energy data from AMI systems. The use of this data, coupled with automated processing, has been identified as the best approach with which to progress M&V (Franconi et al. 2017). The increased accuracy, certainty and standardisation of savings calculations offered by M&V 2.0 is hugely beneficial. To enable this, there is a need to establish guidelines and best practices in order to fully realise the potential of these advancements. This thesis aims to populate this particular aperture in the research area.

M&V, as a sub-sector of the energy industry, is evolving to more dynamic, reactive and accurate methods under the guise of M&V 2.0. This transition brings with it impediments to realising the maximum possible energy savings. Such factors include the increased cost of performing M&V compared to the traditional techniques, maintaining trust in the process when employing black-box modelling techniques, establishing standardised guidance documentation in a developing field and integrating the process deeper into energy management practices. To date, this change has been heavily dominated by the residential and commercial buildings sectors, with the more complex energy systems in existence in industrial buildings delaying the adoption of such advanced techniques. In the industrial buildings sector, the lack of prescriptive guidance on M&V 2.0 and the application of the associated methods are barriers to practitioners adopting such methods. The industry is dominated by proprietary software solutions at present, with many offering only a black-box approach to savings verification. This is a direct contradiction to one of the founding principles of M&V, transparency.

1.2 Research Objectives

The realisation of the energy efficiency opportunity requires successful implementation of demand-side ECMs. M&V plays a critical role in verifying the performance of implemented ECMs to ensure the maximum benefits are being realised. As efforts increase to utilise energy efficiency as a tool in the transition towards low-carbon energy systems, the competence of current M&V methods have come under scrutiny, with current shortcomings in the process being identified as a barrier to financial investment in ECMs. The energy efficiency gap cannot be effectively closed if the challenges impeding M&V from evolving to a dynamic and reactive process continue to exist.

The objective of this thesis is to develop a data science-rooted, accurate, transparent and robust M&V 2.0 solution tailored for the industrial buildings sector. The nucleus of the research addresses the issue of minimising the uncertainty in energy savings, while advancing the process to a near real-time operational state. To achieve this, machine learning is proposed as a powerful tool capable of discovering essential knowledge in the data sets commonly available in industrial facilities. The core research objectives (ROs) of the thesis are defined as follows:

- RO1. Demonstrate the suitability of machine learning techniques to reduce the modelling uncertainty introduced into energy savings quantification with respect to current approaches.
- RO2. Investigate and develop data processing techniques tailored for the domain that ensure efficiency is maintained throughout the process, thus reducing the resources required to carry out M&V.
- RO3. Formalise a prescriptive methodology for the application of machine learning techniques to develop highly accurate baseline energy models for use in M&V 2.0.
- RO4. Integrate the energy modelling methodology into a comprehensive M&V 2.0 framework with a view to embedding performance verification deeper into best practice energy management to ensure persistence of savings over the lifetimes of ECMs.
- RO5. Develop a computationally efficient and intelligent solution for near real-time energy savings quantification and performance deviation detection.

1.3 Scope of Work

The scope of this research is mostly limited to the M&V of energy savings resulting from the implementation of ECMs in the industrial buildings sector. Beyond this, there are conclusions drawn in Chapter 3 relevant to the broader field of demand-side energy modelling. Although the findings of this work are applicable in the residential and commercial buildings sectors, these are not within the scope of the analyses. The complexity of energy systems in industrial facilities require a tailored solution with characteristics that cause it to diverge from residential and commercial buildings. Three case studies were carried out across two large-scale manufacturing facilities in Ireland. In every case, the savings from a real-world ECM were quantified. The challenges faced using real-world data and ECMs are of significance in ensuring applicability of results to solving the challenges facing M&V. In terms of the evaluation of uncertainty in M&V, this thesis focuses on minimising that which is introduced by the baseline energy model. The bounds of this work do not extend beyond the measured data point. Therefore, the uncertainty associated with gathering the data point (i.e measurement) is not considered. However, to ensure a comprehensive approach to the analyses, related literature that addresses measurement and sampling uncertainty is discussed in Chapter 2.

1.4 Outline of Thesis

The remaining chapters of this thesis are outlined as follows (illustrated in Figure 1.6):

- **Chapter 2** presents a comprehensive review of published literature in the research field. The concept of the energy efficiency gap and the motivators and barriers to investment in energy efficiency are discussed. The contribution of performance verification to this phenomenon is then reviewed in the context of energy management practices. The technical challenges facing M&V are highlighted, while potential solutions to the energy modelling issues are discussed in detail. The transition to M&V 2.0 methods is then assessed with a view to identifying the primary solutions needed to ensure the adoption of such approaches. Finally, the field of fault detection and diagnosis (FDD) is leveraged to highlight some key shared learnings that have relevance to the persistence of energy savings over the lifetime of an ECM.

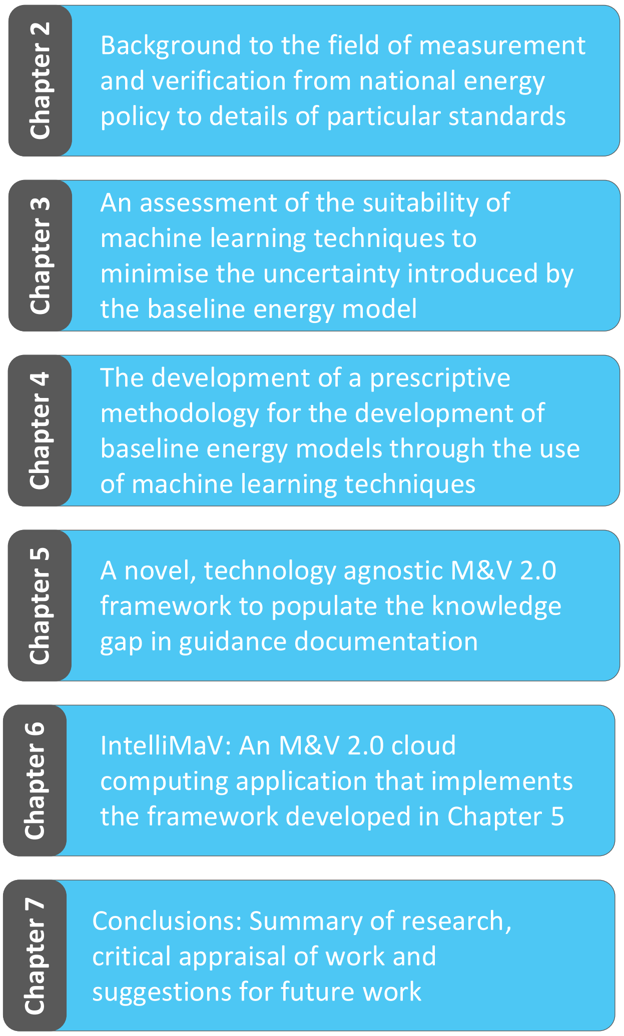


Figure 1.6: Illustration of thesis structure and the contents of each chapter

- **Chapter 3** reviews the suitability of machine learning techniques as a solution capable of minimising the modelling uncertainty in M&V. A case study using real-world data is used to compare the accuracy of traditional energy modelling approaches in M&V to that of the machine learning techniques proposed. A quantitative analysis of both approaches is presented within the context of both the case study and broader research field.

- **Chapter 4** presents a prescriptive methodology for the application of machine learning algorithms in the context of M&V. This populates the knowledge gap that exists amongst the widely accepted M&V guidance documentation on baseline energy model construction. The novel process, which includes data cleaning, feature selection and algorithm application, is presented in an explicit and transparent manner. A case study is again used to quantify the benefits of such an approach under challenging real-world conditions.

- **Chapter 5** consists of an innovative M&V 2.0 framework tailored for the industrial buildings sector. The technology agnostic framework signals a progressive shift away from the proprietary solutions acting as the sole M&V 2.0 resource for industry at present. Particular attention is placed on the persistence of savings through the integration of M&V tasks into the monitoring and targetting (M&T) process as an ongoing energy management solution.
- **Chapter 6** presents a M&V 2.0 cloud computing application that implements the framework detailed in Chapter 5. This fully automated solution for near real-time quantification of savings and performance deviation detection demonstrates the effectiveness of previous research findings and outputs. The application is implemented in a large-scale manufacturing facility with the performance of an ECM tracked in near real-time.
- **Chapter 7** provides a summary of the research contributions in this thesis, which includes a critical appraisal of the work and its limitations. Additionally, suggested directions for future work derived from this research are discussed.

1.5 Research Output

The following publications represent the primary dissemination of the research contained within this thesis to date:

Journal Articles

- **Gallagher, C.V.**, Bruton, K., Leahy, K., O'Sullivan, D.T.J. (2018). The suitability of machine learning to minimise uncertainty in the measurement and verification of energy savings. *Energy and Buildings*, Vol. 158, pages 647-655.
- **Gallagher, C.V.**, Leahy, K., O'Donovan, P., Bruton, K., O'Sullivan, D.T. (2018), Development and application of a machine learning supported methodology for measurement and verification (M&V) 2.0. *Energy and Buildings*, Vol. 167, pages 8-22.

- **Gallagher, C.V.**, Leahy, K., O'Donovan, P., Bruton, K., O'Sullivan, D.T. (2018), IntelliMaV: A cloud computing measurement and verification 2.0 application for automated, near real-time savings quantification and performance deviation detection. *Energy and Buildings*, Vol. 185, pages 26-38.

Conference Proceedings

- **Gallagher, C.V.**, O'Donovan, P., Leahy, K., Bruton, K., O'Sullivan, D.T.J. (2018). From M&V to M&T: An artificial intelligence-based framework for real-time performance verification of demand-side energy savings. In *International Conference on Smart Energy Systems and Technologies (SEST-2018)*, Seville, Spain.
- **Gallagher, C.V.**, Bruton, K., & O'Sullivan, D.T.J. (2016). Utilising the cross industry standard process for data mining to reduce uncertainty in the measurement and verification of energy savings. In *First International Conference on Data Mining and Big Data (DMBD-2016)*, Bali, Indonesia.

Other Publications

In addition to the published research directly related to this thesis, contributions were also made to the following other publications which were led by colleagues in the IERG:

- O'Donovan, P., **Gallagher, C.V.**, Bruton, K., O'Sullivan, D.T.J. (2018). A fog computing industrial cyber-physical system for embedded low-latency machine learning Industry 4.0 applications. *Manufacturing Letters*, Vol. 15(B), pages 139-142.
- Leahy, K., **Gallagher, C.V.**, O'Donovan, P., O'Sullivan, D.T.J. (2018). Cluster analysis of wind turbine alarms for characterising and classifying stoppages. *IET Renewable Power Generation*, Vol. 12(10).
- Leahy, K., **Gallagher, C.V.**, O'Donovan, P., O'Sullivan, D.T.J. (2018). A robust prescriptive framework and performance metric for diagnosing and predicting wind turbine faults based on SCADA and alarms data with case study. *Energies*, Vol. 11(7).
- Leahy, K., **Gallagher, C.V.**, O'Donovan, P., O'Sullivan, D.T.J. (2018). Issues with data quality for wind turbine condition monitoring and reliability analyses. *Energies*, Vol. 12(2).

- Blake, S., **Gallagher, C.V.**, O'Sullivan, D.T.J. (2018). A combined optimization of energy storage system capacity and distributed energy resource scheduling in an industrial microgrid with renewable energy resources. *Sustainable Energy, Grids and Networks* (In Revision).
- Blake, S., **Gallagher, C.V.**, O'Sullivan, D.T.J., O'Donovan, P. (2018). A systematic analysis of microgrid management optimization for microgrids with distributed energy resources. *Sustainable Energy Technologies and Assessments* (In Revision).
- Leahy, K., **Gallagher, C.V.**, O'Donovan, P., O'Sullivan, D.T.J. (2018). Industrial big data pipeline for wind turbine PHM in a large manufacturing facility. *International Journal of Prognostics and Health Management* (Under Review).
- O'Donovan, P., **Gallagher, C.V.**, Leahy, K., Blake, S., Bruton, K., O'Sullivan, D.T.J. (2017). A systematic mapping of industrial cyber-physical system research for Industry 4.0. In *33rd International Manufacturing Conference*, Limerick, Ireland.
- Leahy, K., **Gallagher, C.V.**, Bruton, K., O'Donovan, P., O'Sullivan, D.T.J. (2017). Automatically identifying and predicting unplanned wind turbine stoppages using SCADA and alarm system data: Case study and results. In *Journal of Physics: Conference Series*.

1.6 Novel Contributions

This section outlines the specific novel contributions of this thesis.

Chapter 3 (Gallagher, Bruton, Leahy & O'Sullivan 2018, Gallagher et al. 2016)

- Machine learning techniques are shown to reduce the uncertainty introduced by the baseline energy model when compared to an assumed typical approach to M&V that employs bi-variable ordinary least-squares regression (OLS).
- The use of a higher data measurement frequency reduces the spread of error across the models constructed, within the context of a case study.
- The use of more granular energy data does not always benefit baseline energy model accuracy.

- Sensitivity analysis of the final models shows that those developed using advanced machine learning algorithms on data sets with a higher number of model features can be beneficial in circumstances where missing baseline data limits the model training period length.

Chapter 4 (Gallagher, Leahy, O'Donovan, Bruton & O'Sullivan 2018)

- A novel machine learning supported methodology for M&V 2.0 which enables accurate and reliable quantification of energy savings is presented.
- The methodology includes a novel and computationally efficient feature selection algorithm designed to maintain robustness and trust in the approach.
- Implementation of the methodology in a case study demonstrates its ability to quantify energy savings within acceptable uncertainty limits in challenging circumstances.

Chapter 5 (Gallagher, O'Donovan, Leahy, Bruton & O'Sullivan 2018)

- An innovative, technology agnostic M&V 2.0 framework which utilises the baseline energy modelling methodology developed provides comprehensive guidance in an area devoid of any published literature.
- A performance deviation detection system is incorporated into the framework to ensure persistence of savings over an ECM's lifetime by transitioning M&V into an ongoing energy management activity.

Chapter 6 (Gallagher et al. 2019)

- A cloud computing-based M&V 2.0 application capable of quantifying energy savings with minimal levels of uncertainty is detailed. Advanced machine learning techniques are applied in an automated manner throughout the M&V 2.0 process to leverage the power of the large quantities of energy data that exist in modern manufacturing facilities.
- The use of a cloud computing-based architecture reduces the resources required on-site and decreases the time required to train the baseline energy model through the use of parallel processing.
- Again, a case study is presented using real-world data from a large-scale manufacturing facility to demonstrate the ease of use and benefits of the application.

Chapter 2

Background to Measurement and Verification

2.1 Introduction

This chapter presents a review of published literature in the performance verification of energy savings field. The objective in carrying out this review is to demonstrate the current maturity level of the research and identify knowledge gaps that exist. The scope of this research extends from the role of M&V in closing the energy efficiency gap to the technical aspects of the M&V 2.0 process that are enabling the transition to a mature advanced discipline. Section 2.2 introduces the concept of the energy efficiency gap and the barriers to investment in cost effective ECMs. The current role of M&V in the broader sector of energy management is reviewed in Section 2.3, with a view towards enhancing its presence in the future. Sections 2.4 and 2.5 examine the fundamentals of M&V. Finally, Section 2.6 gives an evaluation of the state of M&V 2.0, while Section 2.7 identifies the lessons that can be learned from the wider field of FDD. Reviews of the machine learning techniques used throughout this body of research are presented in the relevant chapters in which they are employed.

2.2 Closing The Energy Efficiency Gap

As stated in Section 1.1, there exists a significant portion of the energy efficiency opportunity that is not being realised. This difference between the optimal and actual implementation of cost-effective ECMs is known as the energy efficiency gap (Jaffe & Stavins 1994). In recent years, EU policy has sought to close this gap by obligating Member States to achieve 20% energy efficiency savings by 2020. The energy efficiency gap is an area of active research with considerable debate ongoing surrounding the methods used to quantify the magnitude of the gap and the contribution of various factors to its existence. There are two central topics that garner the most attention. These are the methods used to define the optimal level of cost-effective energy efficiency and the most appropriate policy mechanisms for addressing the gap.

2.2.1 Policy Instruments

Developing solutions to these challenges requires a multi-disciplinary approach. For example, different stakeholders possess varying viewpoints when evaluating the optimal level of energy efficiency. There is the economists' economic evaluation, the technologists' assessment, the hypothetical potential, the narrow social optimum and the true social optimum. The quantification of the energy efficiency gap differs amongst each of these evaluations. Analysis has shown that understanding the failures causing this gap and differentiating between the market and non-market failures are required to identify the most appropriate solutions (Jaffe & Stavins 1994).

A market failure is any deviation from the assumptions of a perfect market, which includes economic, organisational and behavioural barriers. Extensive research has been carried out in this regard with energy use externalities and investment inefficiencies being identified as two market failures. It is essential to make the distinction between these two failures when developing policy, as the general theory is that policies should address market failures as directly as possible. It has been concluded that the best policy in cases where both failures exist involves Pigouvian taxes on energy and a mechanism to increase the demand for energy efficient goods and services. A Pigouvian tax is a tax on any market activity that generates negative impacts on externalities that are not directly involved with the activity. A carbon tax is an example of such an approach in which the revenue generated from the additional taxation can be used towards re-mediating the environmental damage caused by carbon emissions. In general, these taxes will cause a reduction in demand which must also be addressed. Examples of mechanisms that increase demand for goods include subsidies and standards (Allcott & Greenstone 2012). To substantiate this finding, government intervention alone has been shown to not always be a viable solution (Klemick & Wolverton 2013).

It is evident that a suite of complementary policy instruments, such as intervention and subsidies, is required to combat negative externalities and investment inefficiencies.

2.2.2 Driving Forces for Investment

There are driving forces stimulating investment in energy efficiency, while barriers simultaneously prevent the full energy efficiency opportunity from being realised. An understanding of both determinants is essential to evolving the industry in a positive manner.

The driving forces, or motivators, for investment in energy efficiency across the industrial sector are most often categorised into four categories: financial, informational, organisational and external. A case study carried out across 65 energy intensive foundries found that financial related driving forces are most the prominent, with organisational ones following suit (Thollander et al. 2013). A deeper review of financial factors found that the threat of rising energy prices and the cost benefits resulting from lower energy consumption were the biggest motivators for investment. Commitment from top management was also found to be an influential factor (Thollander et al. 2013). The results from similar analysis carried out in non-energy intensive industries demonstrated that rising energy prices and the ambition and skill-set of staff within organisations were the two primary driving factors (Rohdin & Thollander 2006).

Within the industrial sector, there has been extensive research undertaken that focused on small and medium-sized enterprises (SMEs), as these are seen as being strategic for competitiveness in increasing energy efficiency. In similar fashion to large enterprises, economic issues, awareness and behavioural issues are the most influential factors in motivating organisations to invest in ECMs (Trianni et al. 2016). Critically, it was found that SMEs experience a large number of barriers and driving forces, which adds to the complexity of achieving the critical aim of improving energy efficiency. Financial driving forces were experienced by 89% of respondents in a case study of 202 SMEs. In the same study, it is interesting to note that a desire to reduce an organisations impact on the environment was only present in 13% of respondents (Meath et al. 2016). Monetary and financial support provided through a variety of policy instruments have been identified as the most appropriate mechanism for increasing SMEs energy efficiency. In a review of manufacturing SMEs, allowances and public financing were the highest ranked driving factors (Cagno & Trianni 2013). Critically, an organisations size has an impact on the perceived driving factors and thus, as stated in Section 2.2.1, a range of instruments must be used to close the energy efficiency gap across the diverse industrial buildings sector.

2.2.3 Barriers to Investments

Table 2.1 highlights the main barriers to investment in energy efficiency in the industrial buildings sector. It is important to note that the research included in Table 2.1 are empirical studies carried out in the industrial sector. Given the nature of this research, the three highest ranked barriers in each study are included for the purposes of this review. Capital constraints, imperfect information and hidden costs are the most prominent barriers to investment identified in this review (Thollander & Ottosson 2008, Rohdin et al. 2007, Meath et al. 2016, Trianni et al. 2016, Rohdin & Thollander 2006, Allcott & Greenstone 2012, Klemick & Wolverton 2013).

The M&V process makes its own contribution to each of these primary barriers to investment. Typically, M&V costs are 1-5% of total project costs when assumptions are employed and 3-10% using a full metering approach (Jayaweera et al. 2013). An industry survey revealed that the cost of energy modelling in M&V can vary based on the desired level of uncertainty (Olinga et al. 2017). This increases the capital required to implement an ECM, thus acting as a barrier to the realisation of energy efficiency savings. Similarly, there is uncertainty associated with any quantification of energy savings, thus adding risk to the project outcome. Therefore, this uncertainty must be minimised to ensure the broad spectrum of ECMs capable of achieving energy savings are implemented. Finally, ambiguity surrounding actual savings achieved contributes to a lack of information, while hidden costs can be a direct result of carrying out M&V without taking a complete approach in considering the affected factors. These issues can be compounded in cases where energy performance contracting (EPC) is utilised as there can be stakeholders from outside the organisation with a financial interest in the outcome of the ECM. This reiterates the need for accurate performance verification.

A compelling topic that is often discussed when evaluating the barriers to investment in cost-effective energy efficiency is the rebound effect. This is the phenomenon that improvements in energy efficiency can induce increases in consumption (Font Vivanco et al. 2016, Sorrell 2007). It has been observed that while the rebound effect needs to be considered by policy-makers, it seems unlikely that all energy efficiency improvements will lead to increases in consumption (Sorrell 2009). On an energy end-use level, the range of estimates for the s ize of the rebound effect is very low to moderate (Greening et al. 2000). The essence of this effect is that all factors directly and indirectly impacted on by an ECM must be considered. This is in line with one of the five fundamental principles of M&V, completeness. A comprehensive performance verification of each and every ECM will contribute substantially to the prevention of the rebound effect.

The current challenges facing M&V are an impedance to closing the energy efficiency gap. The most widely-accepted methodologies are retrospective in their analyses, thus there is potential for savings to deviate from expected levels during the reporting period. In addition, there is a lack of prescriptive guidance on the identification of independent variables and construction of the baseline energy model (Ginestet & Marchio 2010). The costs associated with M&V must also be minimised to ensure the process does not contribute significantly to overall project costs. Beyond the scope of M&V, there is extensive research taking place focusing on the use of advanced machine learning techniques for energy modelling. These practices need to be integrated into the M&V process to advance the tools available to practitioners and populate the information gap that currently exists. This is an active research field with work being undertaken to populate the knowledge gap between previously published literature and the needs of industrial organisations to integrate energy performance in production management (Bunse et al. 2011). This is achievable using a practice known as energy management, which has substantial potential to improve the effectiveness with which energy is managed in a sector that is accountable for such a significant proportion of the world's energy consumption.

Table 2.1: Barriers to investment in industrial energy efficiency as identified in published literature

Barrier	Description (Based on classification defined by Sorrell et al. (2000))	Research
Risk	Risk aversion on a broad spectrum of measures within an organization. For example, the technical risk that production will be disrupted during the implementation of the measure.	Rohdin & Thollander (2006), Klemick & Wolverton (2013)
Uncertainty	There is uncertainty associated with the possible future performance of the ECM.	Thollander & Ottosson (2008), Klemick & Wolverton (2013)
Imperfect information	Stakeholders lack sufficient information to make economically efficient decisions.	Rohdin et al. (2007), Rohdin & Thollander (2006), Trianni et al. (2016), Klemick & Wolverton (2013), Allcott & Greenstone (2012)
Hidden costs	There are additional costs associated with investment in a technology or ECM that are not reflected in engineering models.	Thollander & Ottosson (2008), Rohdin et al. (2007), Rohdin & Thollander (2006), Allcott & Greenstone (2012), Klemick & Wolverton (2013)
Staff constraints	Personnel in an organization lack either the desire, time or skill-set required to improve energy efficiency.	Meath et al. (2016)
Capital constraints	Capital is not available to the end-user to invest in a technology or ECM.	Thollander & Ottosson (2008), Rohdin et al. (2007), Trianni et al. (2016), Meath et al. (2016)
Corporate culture	Energy efficiency is not at the core of the corporate culture of the organisation.	Trianni et al. (2016)
Heterogeneity	An ECM or specific technology may not be cost-effective in all cases within the sector.	Klemick & Wolverton (2013)
Government policy	Government fiscal and regulatory policies, codes and standards.	Langlois-Bertrand et al. (2015)

2.3 Effective Energy Management

It is difficult to establish a single, widely-accepted definition of energy management in published literature, as the interpretation of the term can vary depending on the stakeholder in question. However, an inclusive definition that covers a broad range of use cases is as follows:

"The efficient and effective use of energy to maximise profits and enhance competitive positions" (Capehart et al. 2011)

2.3.1 ISO 50001 Certified Energy Management Systems

It must be acknowledged that financial business decisions are a primary motivating factor in efforts to improve energy efficiency at an end-use level. Both European and National energy policy strive to minimise the human impact on the environment and reduce the carbon emissions; however, businesses most often make decisions to minimise the cost of goods or services delivered. Exceptions to this general case often include compliance with regularity requirements and corporate social responsibility. To this end, energy management systems (EnMSs) are employed to effectively manage energy consumption on the demand-side. An EnMS consists of the establishment of an energy policy and objectives, and processes and procedures to achieve those objectives. The International Organization for Standardization (ISO) have published the most well established standard for establishing, implementing, maintaining and improving an EnMS, ISO 50001 (International Organization for Standardization 2011). The standard is founded on the principle of continual improvement in energy performance and implements the Plan-Do-Check-Act (PDCA) management protocol illustrated in Figure 2.1.

Given the finance oriented nature of the definition for energy management, it is important to note the value of effective energy management to end-users. A methodology for estimating the global, national or regional impacts of ISO 50001 certified EnMSs has been developed and applied under the hypothetical condition that 50% of forecasted global industrial and service sector consumption is accounted for by certified systems by 2030. The value of the energy savings realised through such EnMSs under this assumption was found to be worth $700 billion when discounted to the net present value in 2016 (McKane et al. 2017). These savings are equivalent to 6,500 million metric tonnes of avoided CO_2 emissions. Crucially, this study provides a transparent, consistent process that allows policy makers worldwide to estimate the savings potential from ISO 50001 imple-

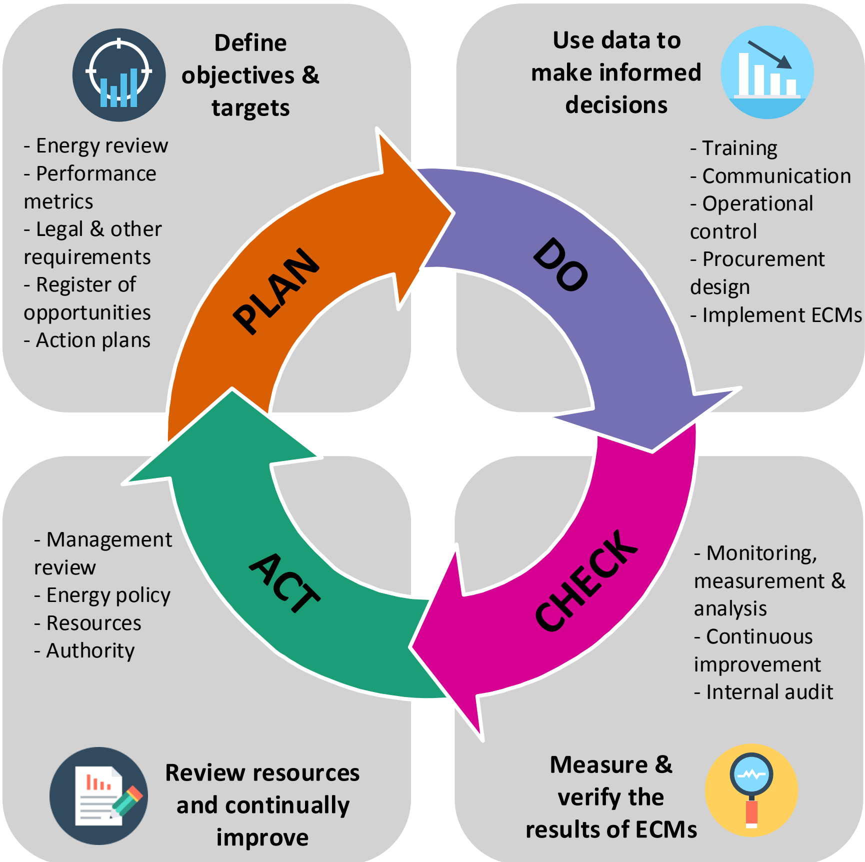


Figure 2.1: Overview of PDCA energy management process

mentation. An empirical review of 57 organisations that adopted ISO 50001 in its early stages found that the greatest motivation for seeking such certification was to improve energy efficiency (Wulandari et al. 2015). With this being an expected finding, the greatest difficulty experienced by these same organisation proves more interesting. The continuous measurement of energy performance was identified as the most laborious task in energy management. This highlights the need for advancements to be made in the manner with which energy consumption and performance are continually measured. This is a task in which M&V has a critical role to play and thus, it must evolve to satisfy the needs of organisations.

2.3.2 Maturity of Energy Management Systems

As a rule of thumb, a properly implemented EnMS typically results in energy savings of 5-10% in the first year of operation with a long-term goal of reducing consumption by 40-50% feasible (Capehart et al. 2011). This highlights the potential influence effective energy management can have on the configuration of energy systems in the future, as maximising efficiency at an end-use level is the first step in solving the challenges presented. It is estimated that ISO 50001 can potentially impact on 60% of the worlds energy use in the future, however, it has previously been concluded that the standard must evolve to tighten controls on energy performance (McKane et al. 2009). This has been considered in an updated version of the standard released in 2018 which places a greater importance on verifying energy performance (International Organization for Standardization 2018).

It is important to note that the ISO do not define any specific requirements surrounding M&V. Clause 4.6.1 of the standard details the requirements for monitoring, measurement and analysis (International Organization for Standardization 2011). These tasks are classified as belonging to the '*Check*' phase of the PDCA cycle. Although the effectiveness of action plans in achieving objectives and targets must be evaluated, the approach to complete this task is at the discretion of organisation implementing the system. Therefore, despite M&V contributing to effective energy management, the technical aspects of such are not detailed and there is not a standard level of performance that must be achieved (i.e. maximum allowable uncertainty for a given confidence interval). Evidence of this can be found in individual implementations of the standard. An example of such can be seen in a tool developed to facilitate the implementation of an ISO 50001 certified EnMS. The *ISO 50001 Analyzer* software has the capability to satisfy the monitoring, measurement and analysis requirements of the standard, but it does not specifically address M&V sufficiently to comply with specific performance standards in this sub-sector of energy management (Gopalakrishnan et al. 2014).

A number of energy management maturity models have been developed to classify the current state of operations of EnMSs and identify the steps required to evolve the system going forward. The growing emphasis placed on M&V across the various levels of maturity defined by these models is an interesting commonality. A maturity model that has been validated in industry defines five levels of maturity (Jovanović & Filipović 2016). Importantly, the final two maturity levels go beyond ISO 50001 compliance requirements and it is at these levels that performance verification must be carried out more comprehensively through the use of statistical analysis. Crucially, in an implementation of the model in nine organisations that operate an EnMS, the monitoring, measurement and analysis practices were found to be on average below these two final levels of maturity. This trait of performance verification becoming more prominent as the maturity of EnMSs evolves is present across published maturity models (Antunes et al.

2014, Ngai et al. 2013, Introna et al. 2014, Prashar 2017, Finnerty et al. 2018, Carbon Trust 2018).

2.3.3 The Superior Energy Performance Programme

It is evident that performance verification will play an increasingly important role in energy management systems of the future. Additionally, it was noted in Section 2.3.2 that the current ISO 50001 standard lacks clear requirements on the quantification of energy savings with respect to uncertainty and confidence levels. The United States (U.S.) Department of Energy (DOE) administer the Superior Energy Performance (SEP) Programme which certifies industrial facilities that have an EnMS that complies with the ISO 50001 standard, while also achieving improved energy performance (US Department of Energy 2018). The SEP Programme places an increased focus on third-party verification of energy performance improvements. This directly improves the reliability and transparency in reported energy savings and places confidence in the M&V process. The DOE defined M&V protocol used to implement the programme compliments the ISO 50001 standard by populating the current gap surrounding the methods used for performance verification.

The use of third-party verification has been shown to enable staff in organisations certified with SEP to more credibly communicate the value of EnMSs to top management, while also demonstrating to other departments in the organisation a willingness to invest in sustainability and reduce production costs (Therkelsen et al. 2013). It is also important to note that SEP participants did not find the increased costs required for third-party verification to be prohibitive. Interestingly, the ratio of savings realised from capital and operational ECMs shifted from 36:64 to 26:74. Savings resulting from operational ECMs often require more advanced M&V methods, which are not a requirement of systems certified to ISO 50001 alone. Thus, there are clear benefits of adopting better performance verification protocols in energy management. Additionally, it has been proven that proper M&V can increase the potential for energy savings under an effective EnMS (Backlund et al. 2012).

2.4 M&V Protocol

2.4.1 Guidance Documentation

There are a number of well established and widely employed methodologies for performing M&V of energy savings. The most prominent of such is the International Performance Measurement and Verification Protocol (IPMVP) published by the Efficiency Valuation Organization (EVO). The IPMVP defines four distinct approaches that can be applied to ensure coverage across the broad spectrum of ECMs (Efficiency Valuation Organization 2016). Options A and B isolate the retrofit with a project boundary encompassing the affected equipment only. Option C differs in that it assesses performance on a whole-facility level and is applicable in cases where the savings are greater than 10% of a facility's total energy consumption. Option D consists of a calibrated simulation of the energy systems and is particularly useful in cases where there is no baseline data available. In similar fashion to the IPMVP, the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE)'s Guideline 14 presents three approaches that mirror the IPMVP options B, C and D. Guideline 14 does not outline any retrofit isolation approach that utilises assumptions for key parameters.

In addition to these comprehensive guidance documents, the ISO has published two standards focused on conducting M&V and determining energy savings at a facility-wide level. ISO 50015 provides general principles and guidelines for the process of M&V of energy performance of an organisation or its components (International Organization for Standardization 2014). The standard can be used either independently or in conjunction with other protocols. ISO 50047 builds upon ISO 50015 by detailing methods for determining energy savings using facility-wide (top-down) and aggregated retrofit isolation (bottom-up) approaches (International Organization for Standardization 2016*b*). These methods can be employed in facilities both with and without EnMSs. Key elements of both ISO standards have been identified and used to develop an "*ISO 50001 M&V Harmonisation Matrix*" (Therkelsen et al. 2018). The objective of the matrix is to enhance credibility in reporting energy performance improvements by establishing a single approach that encompasses the key elements of each individual guidance document. The DOE's SEP discussed in Section 2.3.3 offers yet another set of approaches for verifying the performance of energy efficiency improvement measures (US Department of Energy 2012). The SEP has been applied to the harmonisation matrix as a first step in establishing a single, commonly agreed upon set of approaches to the quantification of energy savings. In addition to the organisation level M&V protocol published by the ISO, ISO 17741 details an ECM level approach, while ISO 17742 focuses on measuring and verifying energy savings at a regional level (International Organization for Standardization 2016*a*, 2015).

In cases where the IPMVP are applied, typical M&V costs are 1-5% of total project costs using Option A and 3-10% using Option B (Jayaweera et al. 2013). It is imperative that any advancements in the approaches to the process do not increase these costs that contribute to the barriers that prevent investment in energy efficiency.

The many different guidance publications available to practitioners are generally based on the same core concepts: accuracy, completeness, conservativeness, consistency, relevance and transparency. The most significant drawback of these protocols is the widely publicised lack of formalised guidance on the development of a baseline energy model, subsequent routine adjustments and estimating the adjusted baseline. Most guidance documents specify that either inverse methods or engineering models can be employed for such a task, but there is little support given to practitioners to properly develop such models in an accurate manner. These provide direction to identify parameters but do not detail a rigid calculation process to follow (Ginestet & Marchio 2010). A lack of such guidance allows these protocol to maintain relevance to the broad spectrum of ECMs, although they are prone to enabling the use of invalid and inaccurate baseline energy models; thus, introducing additional uncertainty to the final savings quantified.

2.4.2 Alternative Approaches

There have been a variety of alternative approaches proposed to expand the knowledge base in the industrial sector and attempt to overcome these aforementioned problems. Data from five industrial buildings was used to compare absolute, intensity and regression approaches to M&V and results showed regression based approaches to be the most effective in translating energy savings values into contextualised energy performance improvement values (Therkelsen et al. 2016).

Elsewhere, a general methodology that takes weather and production into account for measuring plant-wide energy savings has also been developed (Kelly Kissock & Eger 2008). This approach has the ability to disaggregate savings into components which provides additional resolution, while offering a novel, alternative approach to the traditional methodologies. Despite this, the method was found to be limited by the information in the data set, which can be sparse as whole-facility billing data is used.

A prescriptive methodology for performing M&V on combined heat and power (CHP) plants in industrial facilities has been developed (Rossi & Velázquez 2015). In contrast to the protocols reviewed in Section 2.4.1, significant detail is provided on the selection of independent variables and model construction.

A tailored approach to satisfy the requirements of the EC's Energy Performance of Buildings Directive offers an additional alternative to the approaches outlined in the traditional M&V protocol (Burman et al. 2014). This approach details the development of an M&V plan to compare the design and actual energy performance of a building. It was proposed with a view to identifying the shortcomings in the construction and building procurement processes.

2.4.3 Uncertainty Quantification

There are three principle sources of uncertainty in M&V; sampling, measurement and modelling. Similar to the hierarchy of guidance documentation, there exists a number of different approaches for quantifying the uncertainty associated with any quantification of savings. This can directly cause distrust in the savings realised from energy efficiency measures. The IPMVP defines a methodological approach for quantifying each element of uncertainty introduced in a project (Efficiency Valuation Organization 2018). Both EVO and ASHRAE define maximum acceptable levels of uncertainty as the point at which savings are larger than twice the standard error of the baseline model (Efficiency Valuation Organization 2014, American Society of Heating Refrigerating and Air-Conditioning Engineers 2014).

An alternative methodology has also been developed to evaluate the accuracy of building energy models and thus, the uncertainty introduced into a project (Granderson & Price 2014). This approach has been used to evaluate five baseline energy models using data from 29 commercial buildings. It should be noted that this approach uses data from buildings in which no ECM was implemented to quantify uncertainty and offers a useful alternative for comparing the performance of modelling algorithms. The five models evaluated were all found to perform poorly when energy use varied in ways that were not predictable from the outdoor air temperature or the time of week, thus highlighting the challenge faced in industrial buildings in which complex energy systems are impacted by many different variables. This further exposes the need to develop more intricate approaches to improve the applicability to the industrial sector.

All three methodologies highlighted rely on the same core statistical measures for quantifying uncertainty. These are employed to estimate the individual uncertainty introduced by sampling, measurement and modelling. As this thesis focuses on the uncertainty introduced by the baseline energy model, the standard error (SE), otherwise known as the root mean squared error (RMSE), is the primary statistical measure of interest. In it's most simplistic form, the SE can be found as follows:

$$SE = \frac{s}{\sqrt{n}} \tag{2.1}$$

where, n is the number of data points and the standard deviation s is found using the following equation in which y_i is the measured value and $\hat{y}_i$ is the predicted value:

$$s = \sqrt{\frac{\sum_i^n (y_i - \hat{y}_i)^2}{n-1}} \tag{2.2}$$

Finally, the coefficient of variation is utilised to normalise the error metric to the mean. This is expressed using equation 2.3 where $\hat{y}$ is the sample mean.

$$CV = \frac{s}{\hat{y}} \tag{2.3}$$

The core statistical measures detailed are then modified for use in estimating the prediction error of a regression model. The fundamental principle of normalising post-ECM consumption to pre-ECM operating conditions requires the use of a regression model to represent that state of the energy system and estimate the adjusted baseline in the reporting period. Using the IPMVP guidance documentation, the prediction error of the baseline energy model is quantified using Equation 2.4, where k is the number of independent variables in the regression model:

$$SE_{\hat{y}} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-k-1}} \tag{2.4}$$

In line with the trend of difference across the guidance documentation, the ASHRAE Guideline 14 calculates the modelling uncertainty with a slight difference in approach (Equation 2.5). Again, this variance between guidelines causes confusion in the industry.

$$SE_{\hat{y}} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k}} \tag{2.5}$$

In addition to the formula used to compute these error metrics, the methods with which they are applied is also relevant. Previous publications by EVO endorsed the use of all available data in the computation of error metrics (Efficiency Valuation Organization 2014). For example, in the case of modelling error, the model is applied to the entire data set used to train it and thus, the prediction performance metrics are calculated. This is a flawed approach that is prone to over-fitting the model to the available data set, which can result in low levels of model error in the baseline period, but unreliable measures of uncertainty in the reporting period. This approach does not lend itself to the construction of an accurate model generalised to the variation in the data set. The introduction of a random data split overcomes these issues by applying the model to an unseen testing data set. It is important to note that over-fitting only becomes an issue in cases where multiple independent variable are employed. In models with only one or two degrees of freedom, the risk of over-fitting is often negligible.

EVO have subsequently addressed this phenomenon in revised guidance documentation that suggest the use of cross-validation techniques to prevent over-fitting (Efficiency Valuation Organization 2018). Once more, the guidance does not go as far to suggest a specific technique to be applied across the board and establish a common approach. In contrast to this, ASHRAE declare that bootstrapping methods should be used to estimate uncertainty. Cross-validation and bootstrapping differ in their approaches with each posing unique advantages and disadvantages.

The most common cross-validation technique used is k-fold cross-validation, which is often called rotation estimation. Using this technique, the data set is randomly split into k mutually exclusive subsets (or folds) of approximately equal size. The model is trained and tested k times with the testing data set varying each time. Figure 2.2 illustrates the process involved. Leave-one-out cross-validation is a more expensive technique in which k equals the number of data points, n. EVO does not specify which technique should be employed; however, ten-fold cross validation has been shown to be better than the more expensive leave-one-out approach (Kohavi 1995). Cross-validation tends to be less biased than bootstrapping, but often results in higher variance.

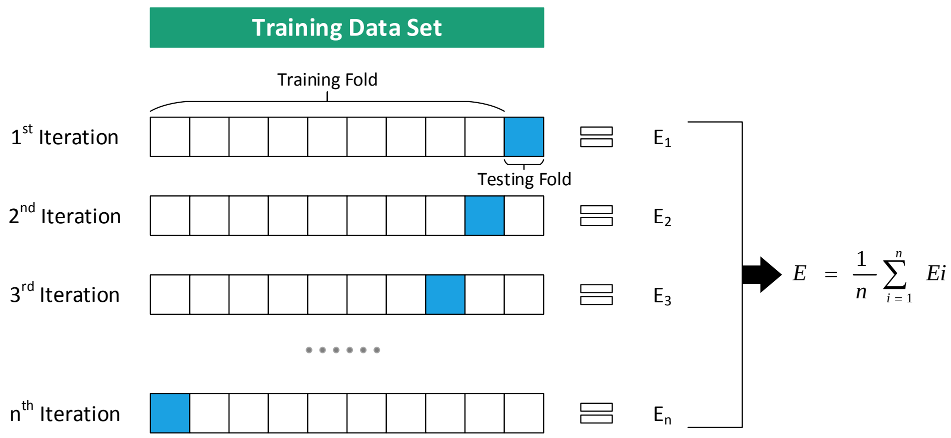


Figure 2.2: Illustration of k-fold cross-validation process

With ASHRAE promoting the use of bootstrapping, it is important to identify the differences between this approach and that endorsed by EVO. Introduced in the late twentieth century, bootstrapping is a statistical method for estimating the sampling distribution of an estimator by sampling with a replacement from the original sample (Efron & Tibshirani 1993). Figure 2.3 illustrates the bootstrapping process. The principle is that each sample is representative of the original population. Bootstrapping has the potential to reduce variance as a distribution of the desired statistical measure is calculated, but this comes with a trade-off in bias. A combination approach is often preferable in which cross-validation is used to select the regression model and bootstrapping is used to quantify the uncertainty associated with the final model. Thus, the EVO and ASHRAE approaches are complimentary to each other; however, a framework for their harmonised application is required.

The impact of uncertainty beyond the scope of a single ECM has been discussed in Sections 2.2 and 2.3. In this section, the discrepancies in guidance given to quantify modelling uncertainty between the established protocols has been highlighted. These discrepancies contribute to investors having a lack of trust and confidence in the savings realised from energy efficiency measures. In line with the core concepts of M&V, consistency amongst the approaches taken to quantify savings and reducing uncertainty in individual projects is critical in enhancing the standing of energy efficiency as a reliable tool capable of delivering results in an accurate manner. Correcting the inefficiencies at a project level will also contribute to the solutions needed to overcome the current barriers to

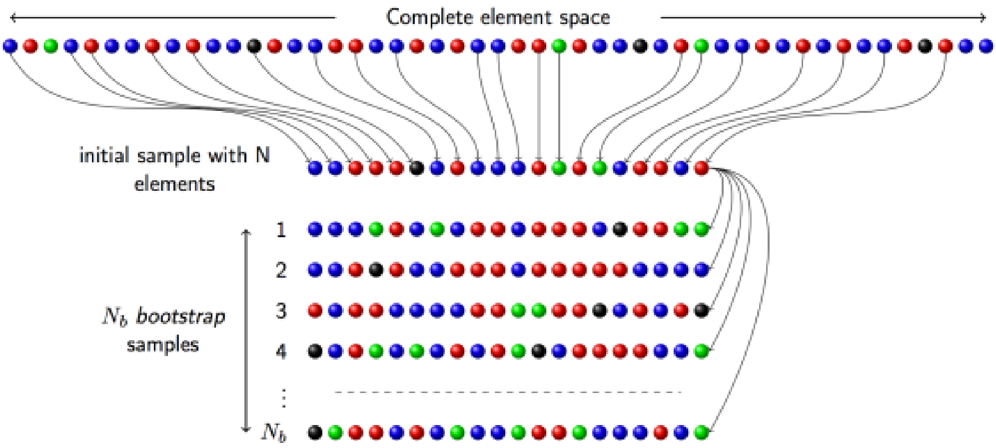


Figure 2.3: Illustration of bootstrapping process, reprinted from Seth (2017)

investment at a macro-level.

2.5 Energy Modelling in M&V

The critical step in any M&V methodology is the development of an energy model prior to the implementation of the ECM. This is referred to as the baseline energy model. The approaches taken to model this baseline energy consumption need to evolve in order to fully capture the behaviour of complex energy systems in industrial buildings. The use of more advanced algorithms, beyond that of the typical OLS techniques commonly applied, enables practitioners to accurately model the energy consumption behaviour in these scenarios; hence improving the certainty with which savings are quantified.

As stated in earlier sections, the research presented in this thesis focuses on minimising the uncertainty introduced by the baseline energy model employed. Developing an understanding of the modelling process is key to working towards a solution that maximises the accuracy of each baseline energy model. System identification addresses the problem of constructing mathematical models of dynamic systems based on observed data of the system's input and outputs (Ljung 1987). It is best practice in system identification to utilise prior knowledge and physical insight about the system, when available, in selecting the model structure. In cases where this information is not available, the relationship between inputs and outputs must be derived without any understanding of the internal workings of the system. The different types of model that can be applied for modelling energy consumption in buildings are generally classified into three categories; white-box, grey-box and black-box (Sjöberg et al. 1995). Figure 2.4 provides a high-level comparison of the three types of models. The generalised definitions for each are included below.

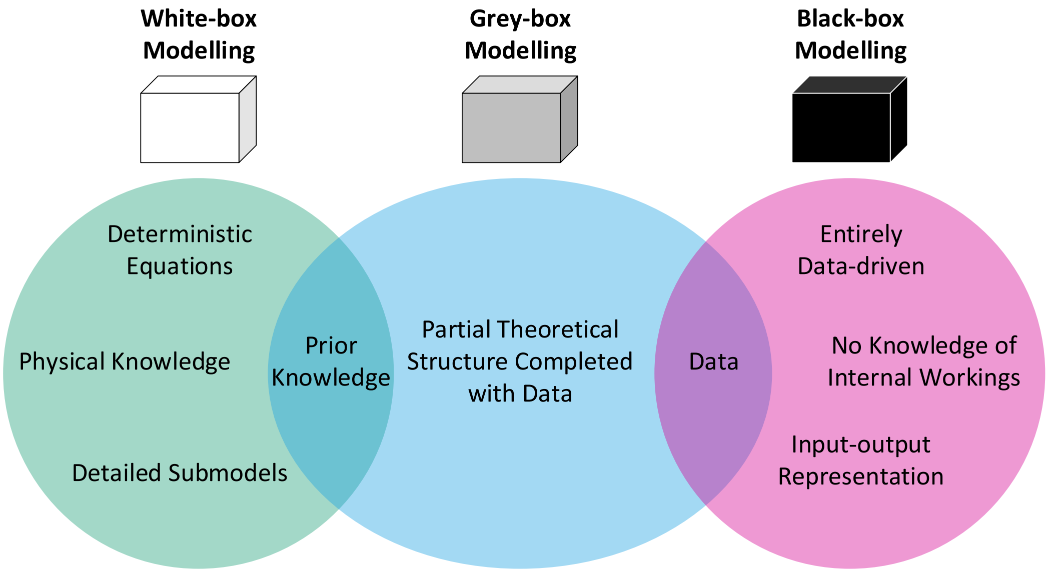


Figure 2.4: Comparison of modelling classifications

- *White-box models:* The model is perfectly known and can be constructed entirely from prior knowledge and physical insight (Sjöberg et al. 1995).
- *Grey-box models:* Uses some theoretical structure that is not complete and thus, the model is completed using data (Whiten 2013). They are subdivided into two classes:
  - *Physical models:* Physical knowledge can be used to build a model, but a certain number of parameters must be estimated from the available data (Sjöberg et al. 1995).
  - *Semi-physical models:* Physical knowledge is used to infer some non-linear combinations of measured data signals. These inferences are then subjected to black-box modelling techniques (Sjöberg et al. 1995).
- *Black-box models:* No physical insight is available, thus the model is usually formed from a known basis of model terms which is sufficiently flexible to represent a broad range of behaviours (Worden et al. 2007).

Figure 2.5 illustrates the useful process flow of a building energy system to be modelled. The process consists of three stages which represent the system and the factors affecting it. Controllable variables include anything that can be controlled by system users, while uncontrollable variables are external input variables such as outdoor air temperature. The availability of information for each stage dictates which modelling approaches are applicable in individual cases. For example, in cases where the input variables and the structure of the energy system model are known, the outputs can be estimated using a physics-based (or white-box) model that does not rely on measured data. This is only useful in cases where the energy transfer processes are known and thus, it can be a resource intensive task in circumstances of complex energy systems.

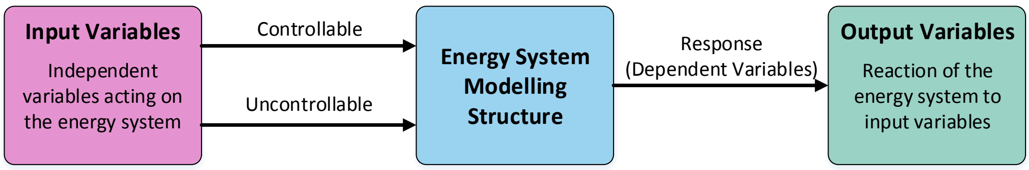


Figure 2.5: Process flow diagram of regression modelling task (Harish & Kumar 2016)

As discussed in Section 2.4, there is no prescriptive guidance on the construction of the baseline energy model in the established protocols. This has led to a wide-variety of approaches being applied with varying degrees of success. The following section reviews these applications with the objective of identifying the most appropriate solution for energy modelling in the context of M&V. As the task is not exclusive to performance verification, it is also prudent to include applications in the broader field of building energy modelling.

The difference between energy modelling in residential, commercial and industrial facilities is also noteworthy. The techniques used in residential and commercial buildings are similar as the primary factors that impact on energy consumption are often the same. These include occupancy, outside air temperature and building characteristics. In contrast, industrial buildings contain multiple factors that affect the more complex consumption and the savings realised are often small relative to the total facility consumption. The available information stored within the vast quantities of data that are common in industrial facilities offers a powerful opportunity to advance the subject area.

The standard regression techniques utilised in residential and commercial buildings have been applied to billing data for industrial facilities (Golden et al. 2017). The three parameter cooling and cooling degree days (CDD) models were applied to real-world data for two industrial buildings. In these cases, the approaches were capable of meeting the ASHRAE recommended values for the cofficient of variation of the root mean squared error (CV(RMSE)) and normalised mean bias error (NMBE), although the coefficient of determination (R^2) values in both instances suggested poor correlations between modelled and measured values. This suggests an over-reliance on the ASHRAE recommendations that fail to account for R^2 in their approach. The importance of a statistically valid R^2 (usually greater than 0.7) is particularly important when monthly billing data is used as information is sparse. The use of these residential and commercial buildings methods is also reliant on estimations of the energy consumption of critical systems, such as production and heating, ventilation and air-conditioning (HVAC), for disaggregating whole-building consumption. Estimations are typically not useful in M&V as the use of assumptions must be limited to ensure accuracy. This case study provides a useful example of the issues encountered when solutions developed for the residential and commercial buildings sectors are applied to industrial buildings.

2.5.1 Interdisciplinary Reviews

As energy modelling in buildings is quite a mature field, there have been a number of extensive reviews carried out that assess the effectiveness of numerous modelling approaches. A comprehensive review of models developed using simplified engineering, statistical and artificial intelligence (AI) methods for the modelling and prediction of energy consumption in buildings concluded that the machine learning artificial neural network (ANN) and support vector machine (SVM) techniques are the most widely used models (Zhao & Magoulès 2012). This study contains many relevant findings for assessing the state of end-use energy system modelling. The elaborate nature of engineering models developed using a white-box approach were found to be accurate in application and capable of becoming lightweight by adopting simplifying strategies. The critical drawback of such methods was identified as being the difficulty of execution due to high complexity and a lack of input information. This issue is exasperated in large industrial facilities in which multiple variables impact on the performance of the energy system.

Black-box approaches were analysed in detail with both simple statistical and advanced AI techniques reviewed. The simplicity of the statistical models reviewed is advantageous, although the inaccuracy and lack of flexibility of such approaches limits their use. ANNs and SVMs were found to perform well in modelling non-linear relationships, hence, making them very applicable for energy consumption prediction. SVMs have been proven to perform better than ANNs when predicting consumption in residential buildings (Li et al. 2010). Critically, it is difficult to identify any one modelling approach that is most appropriate for all applications in building energy modelling. An exhaustive comparison under a set of common conditions is necessary to be able to conclude on such a hypothesis (Zhao & Magoulès 2012). Additionally, research findings suggest more application guidance is needed to fully empower end-users in the application of advanced machine learning techniques (Yildiz et al. 2017).

When white, grey and black-box modelling approaches are directly compared, it has been found that black-box models are the easiest to deploy and grey-box models offer significant promise in the future (Foucquier et al. 2013). The data-driven black-box approach is also capable of reducing the time taken to create an energy system model, while maintaining an adequate level of predictive accuracy (Tardioli et al. 2015). ANNs and SVMs were reviewed with the aim of identifying better forecasting of building's electricity consumption and it was concluded that they are advantageous in capturing the behaviour of complex energy systems influenced by many parameters (Ahmad et al. 2014). This research substantiates findings from other works that ANNs and SVMs are the most widely used machine learning techniques in building energy system modelling. In the area of AI-based black-box models, ensemble models have been identified as the most accurate and stable approach when compared to ANNs, SVMs and OLS, however, the low computational speed and difficulty are a significant hindrance to the their adoption at present. The application needs to be simplified to effectively employ AI-based methods in the real world (Wang & Srinivasan 2017).

2.5.2 Case Study Applications

White-box Approaches

It would be insular to review energy modelling within the sole context of M&V, as there are many other areas which require a similar task to be completed with accuracy and precision. Therefore, it is prudent to identify important findings on demand-side energy modelling in an effort to advance the research field of interest. White-box or physical models can be classified into three categories depending on the approach employed. These are the computational fluid dynamics (CFD), zonal and nodal approaches (Foucquier et al. 2013).

The CFD method is viewed as the most complete approach to thermal building simulation. It consists of a granular analysis of the thermal transfer modelling that allows the flow field to be detailed. Each building zone is divided using a large number of control volumes with a homogeneous or heterogeneous global mesh. It is only practical to apply the CFD approach using simulation software and the large computation time required is a significant hindrance to its use. It has been found that a coupled approach of building simulation and CFD simulation is capable of improving the accuracy and efficiency in natural ventilation prediction (Wang & Wong 2008). Despite this, the model complexity, skill required and computation time result in the CFD approach being inappropriate for application in performance verification.

Zonal methods simplify the CFD approach by dividing buildings into zones and subsequently, partitioning these zones into several cells. A zonal model has been used to predict the air temperature distribution inside a room (Inard et al. 1996). The detail provided from such models are generally not required within the scope of M&V. As with the CFD, the computation time required to solve the problem is a prohibiting factor.

The nodal approach is the simplest white-box method for energy modelling. Often called the multi-zone technique, a node generally represents a room, wall or the exterior of the building. However, there is scope to be more specific in defining each node. Fundamentally, one node represents a building zone that is characterised by uniform state variables. There are many software applications that utilise the nodal approach. The U.S. DOE maintains a database of the tools available for applying such methods (US Department of Energy 2017). One such tool is *EnergyPlus* software which was used to create a model of a commercial building's thermal behaviour. The high accuracy of the model was validated by comparing simulated and measured data for temperatures throughout the building (Royer et al. 2013).

The use of measured data to improve the prediction performance of simulation models is referred to as calibrated simulation, which is defined as Option D in the IPMVP. The benefits of such an approach include increased confidence in the M&V process. However, a calibrated simulation approach requires expert skill, system knowledge and significant resources to complete. A detailed review of these approaches was critically assessed in published literature and provides an insight into the field (Reddy 2006). The prohibitive model complexity has been reduced while maintaining accuracy through the use of Laplace transforms, physical and generalised parameters, and measured data (Lü et al. 2014).

In another simplified approach, daily energy consumption has been modelled for the design of a renewable energy system in residential buildings. This was achieved by constructing the model based on minimal input parameters such as daily energy consumption of appliances, ownership of each appliance and occupancy schedules (Yao & Steemers 2005). In a comparison of detailed simulation and simplified equivalent models for estimating cooling load profiles, there was higher variance in the simplified approach when individual buildings were assessed. However, there was little bias in the grouped results. The detailed models are more beneficial in circumstances where more granular detail is required (Yik et al. 2001). It has been highlighted that the quantification of uncertainty in the specification of model parameters is often neglected, which creates ambiguity surrounding the accuracy of the model outputs (Coakley et al. 2014).

Grey-box Approaches

Grey-box (or hybrid) modelling consists of firstly formulating a physics-based model to represent the system using white-box techniques. Subsequently, important parameters that are representative of certain key and aggregated physical parameters and characteristics are identified using statistical analysis (Rabl & Rialhe 1992). This process requires a high-level of expertise in establishing the physics-based model and estimating the parameters. It has been identified as being particularly advantageous for FDD, although its applicability to whole-facility energy use modelling is limited (Harish & Kumar 2016).

In most cases, physics-based models and measured data are combined to enable the development of an accurate regression model. There are many examples of this approach being employed to accurately model the behaviours of energy systems in buildings. Hybrid models are at their most useful when a physics-based model of the building is available, but is incomplete or does not offer enough detail and thus, it must be adapted. This solves the difficult task of rebuilding white-box models in existing buildings (Foucquier et al. 2013). They offer an alternative to black-box models, which require large amounts of training data, by developing a simplified building thermal model and method to identify the parameters of the model using operational data and genetic algorithms (GAs). Results in a commercial office building demonstrated the ability of the model to perform with robustness and accuracy under different operating conditions (Wang & Xu 2006).

In early work, a grey model was utilised to predict the consumption of a building humidification system with reasonably high accuracy (Wang et al. 1999). Similarly, grey system theory has been used to predict the energy consumption of a domestic air source heat pump water heater (Guo et al. 2011). The prediction performance of the model improved with increases in the sample interval. Thus, more granular data did not aid model accuracy. The small quantity of measured data required by the model is noteworthy.

In an application that demonstrates the benefits of grey-box modelling for short-term predictions, measured data for climatic conditions were combined with a grey model to predict the next day hourly thermal load of buildings (Zhou et al. 2008). The integration of weather data improved the prediction performance of the thermal model. Additionally, a PID controller, fuzzy logic and a dynamic model were combined for use in an indoor temperature controller in a building by describing its thermal behaviour (Paris et al. 2010, 2011).

The use of the now widely-employed ANN in combination with a physics-based thermal model has been proven to improve the efficiency in HVAC systems (Teeter & Chow 1998). This approach reduces the complexity and speed of model training. Also, utilising an ANN for identification and control provides a means of adapting a controller in-situ in an effort to minimise cost. In a large geographical area, a physical model for heat consumption considered together with measurements of climatic conditions has proven to be powerful. Heat transfer theory was used to create an initial model structure and data on heat consumption and weather was utilised to establish a mathematical model of the heat consumption (Nielsen & Madsen 2006). Commonly-used building energy model simulation tools have also been combined with measured data using a hybrid approach. The *EnergyPlus* software and an ANN were used to predict the energy consumption in a network of buildings. The simulation time was drastically reduced from approximately one hour to seconds using the hybrid-approach (Xu et al. 2012).

Black-box Approaches

Black-box models are often referred to as *"data-driven models"*. The advent of AMI has resulted in the availability of large quantities of granular energy data and black-box modelling is capable of exploiting the knowledge contained within these data. The process consists of using single or multi-variable regression analysis to map measured outputs to measured inputs of the energy system. AI is a growing area of interest that is becoming more prominent in the building energy modelling field due to the power of its techniques. The benefits of such an approach include the requirement for less measurement data, reduced model construction time and computational efficiency (Harish & Kumar 2016). AI has been proven to be advantageous in building energy load prediction (Wang & Srinivasan 2017). The term AI is inclusive of machine learning techniques, which are a subset of AI.

A data-driven approach to modelling is particularly advantageous in complex industrial facilities where prior knowledge of the physical relationships within the building are unknown. In these cases, constructing white-box models can be a very complex and resource intensive task. Given the nature of single use models in M&V, the development of such a physics-based model would be too costly under the project constraints. Therefore, black-box models offer a low cost, yet powerful alternative that is capable of discovering and utilising the knowledge within the vast quantities of data collected by AMI in modern industrial facilities.

An important characteristic of these models is the training data required to accurately construct them. The model training period and the measurement frequency of the data are variable and have been found to significantly influence prediction performance. The effect of measurement frequency on the performance of temperature dependent regression models was reviewed using training data varying from 1 day to 3 months. Results show that the relative error for predicting annual energy consumption was 100% and 6% respectively (Chou & Bui 2014). Additionally, hourly data was identified as the most appropriate for multi-family residential applications when measurement frequency varied from 10-minute to daily intervals (Jain et al. 2014). Crucially, a review focused on prediction scopes, the properties of input data and pre-processing methods concluded that there is no one-size-fits-all model that can be utilised under all conditions (Amasyali & El-Gohary 2018).

There are several examples of machine learning techniques being used to successfully predict energy consumption and related factors in buildings. Both energy consumption and local environmental conditions have been employed to complete a multi-objective optimization of a HVAC system performance. This resulted in a multi-layer perceptron (a class of feed-forward ANN) ensemble approach accurately modelling the systems state (Wei et al. 2015, Tang et al. 2014).

On a less granular scale, a simplified method for estimating hourly energy consumption of a building by applying a series of predetermined coefficients to the monthly energy consumption data from utility bills is proposed (Fumo et al. 2010). Billing data is a high-level source of information that is widely available to every building operator. Gaussian process (GP) and three-parameter cooling change point models have been developed to utilise this billing data for energy consumption modelling (Carpenter et al. 2016, Golden et al. 2017).

In line with their common use across sectors, SVMs have been used to forecast the energy consumption of four commercial buildings using outdoor air temperature, relative humidity and global solar irradiation as predictor variables (Dong et al. 2005). In the commercial buildings sector, SVMs have been extensively proven to be accurate in modelling energy consumption (Li et al. 2009*a*,*b*, Dong et al. 2005, Edwards et al. 2012).

Similarly, ANNs were used to accurately model the baseline energy consumption of CHP plants (Rossi et al. 2014). Substantiating findings in the literature review studies discussed earlier, there are a wide-range of studies proving the suitability of ANNs for modelling buildings energy consumption (Wong et al. 2010, Karatasou et al. 2006, Haberl & Thamilseran 1996, Kreider et al. 1995). In the early 1990's, the ASHRAE Great Energy Predictor Shoot-out identified ANNs as the most accurate method of modelling a building's energy use (Kreider & Haberl 1994*b*,*a*). In the second ASHRAE Great Energy Predictor Shoot-out two years later, hourly whole-building data was used by four competitors to model the energy consumption of commercial buildings. A machine learning approach employing ANNs won the competition, although a statistical regression method was found to perform almost as well as the ANNs (Haberl & Thamilseran 1996).

A comparison between physics-based modelling and ANNs for forecasting energy consumption highlights the requirement for training data as a hindrance to using the ANN, despite it performing as well as the white-box model (Neto & Fiorelli 2008). This requirement is common across all machine learning algorithms and the performance of the models often depend on the quantity of training data available. However, research in more recent years focusing on grey-box techniques has been conducted on the use of physics-based modelling to train machine learning models and this offers further potential for the utilisation of machine learning in M&V (de Wilde et al. 2013).

Energy data with an hourly measurement frequency were used to show that the deep recurrent ANN performs better than the multi-layer perceptron ANN for electricity forecasting in residential and commercial buildings (Rahman et al. 2018). Machine learning models trained using a national data set for commercial buildings can be applied to predict energy consumption using only a small number of features such as square footage, building activity, heating degree days (HDD), CDD and the number of floors (Robinson et al. 2017).

Piecewise linear regression and GP models were compared for predicting energy consumption in industrial buildings with both performing similarly well and meeting ASHRAE Guideline 14 minimum requirements (Carpenter et al. 2018). Predictive control based on historical building data has been achieved by utilising machine learning algorithms such as regression trees and random forests (Smarra et al. 2018). The gradient boosting machine learning algorithm was assessed relative to models trained using piecewise linear regression and random forest algorithms to model energy consumption in 410 commercial buildings (Touzani et al. 2018). The gradient boosting machine model improved the prediction accuracy in 80% of cases.

In the residential buildings sector, the available data for developing baseline models is often restricted to whole building consumption, outside air temperature and occupancy. It is important to note that despite occupancy being highly correlated with energy use, it does not always significantly improve the accuracy of the baseline model (Liang et al. 2016). In commercial buildings, the orientation, insulation thickness and transparency ratio can be used to develop an ANN capable of predicting heating energy consumption with 94.8-98.5% accuracy (Ekici & Aksoy 2009). Similarly, a building energy demand predictive model based on the decision tree method performed with an accuracy of 92% on a test data set (Yu et al. 2010).

In the context of M&V, GP modelling was used to determine energy savings and uncertainty levels in commercial office buildings. The models developed were capable of capturing the complex non-linear and multi-variable interactions, as well as multi-resolution trends of energy behaviour (Heo & Zavala 2012). Inverse simulation has been proposed as a less time-intensive method of estimating energy savings in the industrial sector. Using this approach, savings can be determined using a multi-variable three-parameter change-point regression model driven with typical weather data (Sever et al. 2011). In a review of 10 baseline energy models used for whole-building M&V, the algorithms used to train models included principle component analysis (PCA), random forests, mean-week and time approaches, advanced regression and k-nearest neighbours (k-NN) (Granderson et al. 2016).

A methodology for measuring whole-facility industrial energy savings that accounts for weather and production can use sub-metered data or whole-plant utility billing data (Kelly Kissock & Eger 2008). The purpose of the methodology is to extract information about savings from the data set; however, this is limited by the quality of the data set itself. The use of monthly data was noted as a significant limiting factor in analysis of ECMs on different time-scales. This research highlights the potential benefits of employing granular energy data for the purposes of M&V. Additionally, a prescriptive methodology for performing M&V on CHP plants in industrial buildings has been developed (Rossi & Velázquez 2015).

Deep learning methods have also been applied for the purposes of modelling demand-side energy consumption with a method for accurately predicting consumption 24-hours in advance published (Fan et al. 2017). This research found that non-linear techniques performed most accurately, with an extreme gradient boosting method the most accurate model. It also found that supervised deep learning techniques did not show evident advantages in developing cooling load prediction models with a 'shallow' architecture performing as well as its deep counterpart.

These novel M&V methodologies are beneficial to progressing the research field, although they are only applicable under specific conditions. Research has established that data-driven energy modelling approaches are capable of performing better than traditional approaches, while requiring significantly less input data from the end user (Edwards et al. 2012, Zhao & Magoulès 2012). In this section, the research published to date is reviewed to depict the state of maturity in each sector. Within the scope of M&V, the majority of research published to date focuses on residential and commercial buildings. This has led to a knowledge gap surrounding the application of these complex algorithms in industrial facilities for the purposes of energy savings verification. The lack of research investigating energy modelling in industrial M&V applications, coupled with the success achieved in residential and commercial applications, are strong indicators of the potential advancements that are possible. The M&V methods across all applications hold many commonalities; however, without specific methodologies that address the requirements of each case, the accuracy of energy savings estimation is restricted. The methodology proposed in Chapter 4 satisfies this need and enables M&V in industrial facilities to progress to a more dynamic and reactive state (M&V 2.0).

2.5.3 Measurement and Sampling Uncertainty

It was stated in Chapter 1 (Section 1.3) that the research detailed in this thesis focuses on minimising the uncertainty in M&V introduced by the baseline energy model. Measurement and sampling uncertainty do not lie within the scope of this thesis. Despite this, it is prudent to highlight the prominent literatures related to measurement and sampling uncertainty. It is also important to note that the use of the more advanced modelling techniques already discussed can often rely on larger quantities of data from a greater number of variables. This will be investigated in detail in Chapter 3, however, an unintentional increase in measurement uncertainty is a consideration when striving to reduce modelling uncertainty. This interactive affect is caused by the utilisation of a greater number of variables to construct the baseline energy model.

It has been recommended that sensitivity analysis should be carried out to investigate the affects of errors in measurement on baseline energy models (Carstens et al. 2018*b*). If errors are found in data recording methods, a measurement error model is necessary to compensate for bias introduced. A low-cost method for calibrating energy meters, that utilises machine learning tools in a hybrid approach, offers a solution to measurement uncertainty issues (Carstens et al. 2017*b*).

The costs and labour hours required to carry out performance verification are limited in any project, as M&V is an overhead cost. In an attempt to identify the optimal method for allocating resources, a cost optimisation model has been developed to account for sampling and modelling uncertainties (Olinga et al. 2017). Within the scope of lighting projects, a metering cost optimisation model has been proposed to minimise metering costs, while satisfying accuracy requirements (Ye & Xia 2016). Finally, an approach has been developed that combines and quantifies all three sources of uncertainty in M&V by utilising Bayesian statistics (Carstens et al. 2017*a*).

2.5.4 Integration of Advanced Modelling Algorithms to M&V

The ability to integrate the advanced machine learning algorithms discussed previously into the M&V process offers the benefit of improved prediction performance, thus leading to less uncertainty in savings quantified. A key criteria of many of these algorithms is their dependency on large quantities of training data spread across multiple variables to enable the discovery of sufficient knowledge on the quantity being modelled. This requires a data pipeline to integrate the M&V application with data readily available in industrial facilities. A fog computing cyber-physical system for embedding low-latency machine learning applications in an Industry 4.0 environment represents an example of a solution to data integration (O'Donovan et al. 2018). Similarly, a computational framework that enables remote real-time monitoring and scalable high-performance computing applications through the utilisation of wireless sensor networks, cloud computing and machine learning has been developed with this in mind (Wu et al. 2017). These can all be used to enable the advancement of M&V in the modern industrial facility.

In addition to the data pipelines required to realise the true power of such algorithms, prescriptive guidance on the steps required for their accurate and valid application is also required. This has been highlighted as a problematic shortcoming of the current widely-accepted M&V protocols. Success in this regard will empower practitioners, who do not necessarily possess subject matter expert knowledge on the algorithms, to apply powerful data-driven knowledge discovery techniques for the betterment of performance verification.

2.6 Measurement and Verification 2.0

It has already been highlighted in Chapter 1 that M&V practices are undergoing a transition to a state most commonly referred to as M&V 2.0. This has been defined as the use of automated analytics on more granular data sets (in terms of frequency, volume and end-use detail) to perform continuous, near real-time quantification of energy savings (Franconi et al. 2017). This evolution of practices is resulting in increases in accuracy and effectiveness which are made possible by combining more detailed data sets with automated processing of analytics. One of the key benefits of M&V 2.0 is the labour hours that can be saved relative to traditional, more-simplistic approaches, while still meeting uncertainty criteria (Granderson et al. 2017).

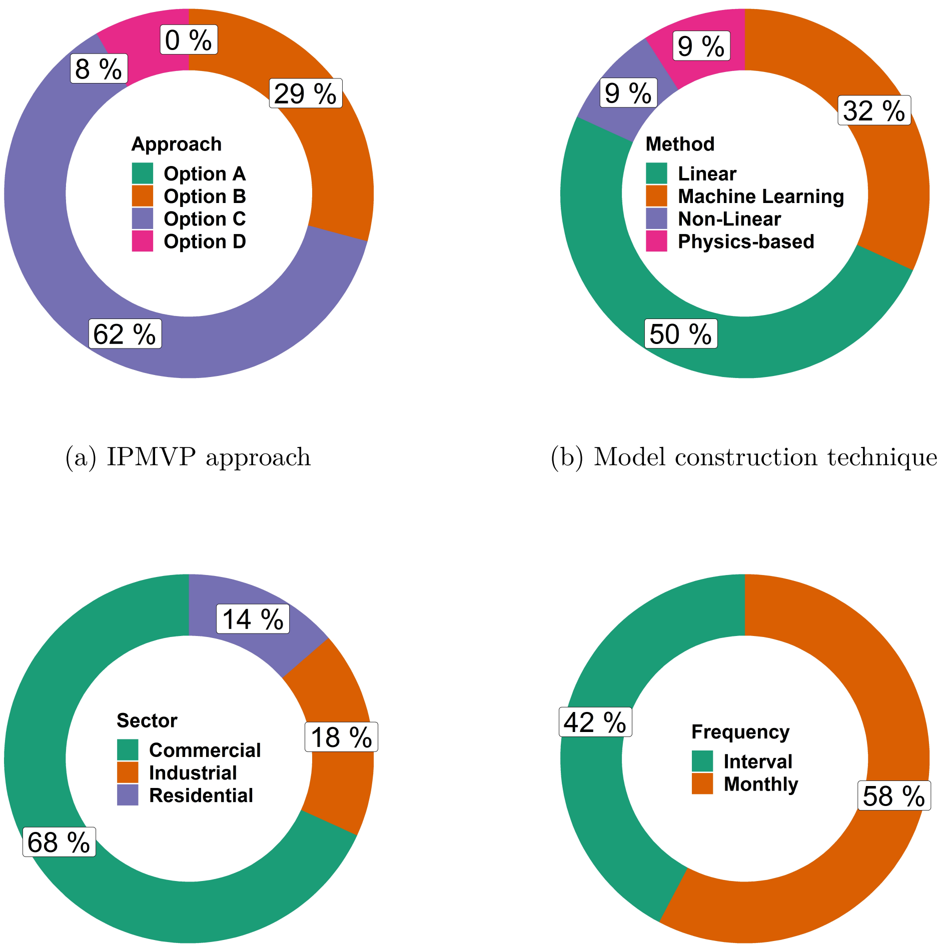
There are a number of factors that that are driving this evolution of M&V practices. The availability of large-quantities of granular energy data recorded by AMI, which is now common place in industrial buildings, is one such factor. In addition, real-time monitoring and utilising advanced statistical models for forecasting have been highlighted as key characteristics in a mature and optimised EnMS (Jovanović & Filipović 2016). EPC is also becoming more prominent in the energy services industry and the foundations of the entire practice relies upon accurate, transparent and trustworthy M&V. Uncertainty has been identified as a key risk to energy services companys (ESCOs) in performance contracting (Lee et al. 2015). An investigation into the drivers of digitalisation and Industry 4.0 in the coal and steel industry concluded that although energy efficiency is not the core driver, it should be positively affected by these movements with the concept of zero-waste-production leading to an increase in energy efficiency (Arens et al. 2018). It was also found that the word 'energy' features in the title or abstract of 10.6% of the projects associated with the research fund for coal and steel.

A detailed review of solutions with M&V 2.0 capabilities available on the market covered sixteen offerings (Granderson & Fernandes 2017). Despite this not being an exhaustive review as the market changes frequently, it does give a good indication of the current level of maturity of M&V technology. Of the sixteen technologies examined, four were applicable to the industrial sector and of this subset, the modelling characteristics varied from OLS to more advanced machine learning techniques. Figure 2.6 illustrates the findings from this study and offers a useful insight into the current market state. An earlier review that describes the vendor landscape for M&V 2.0 offers focused on the residential and commercial buildings sectors and found that the industry is still in its relative infancy (Kupser et al. 2016).

It can be difficult and problematic to apply the same techniques to industrial, commercial and residential buildings as the data available to M&V practitioners in each sector is different. Energy systems in the residential and commercial buildings sectors are most often affected by the similar independent variables. These include occupancy, system scheduling and outside air temperature. This allows for similar methods to be used in both cases. However, the complex energy systems that exist in the industrial buildings sector presents a problem in that many factors, some of which are unknown, impact on energy consumption. Thus, it is prudent to review the sector specific M&V 2.0 applications and solutions in isolation.

2.6.1 Residential and Commercial Applications

To date, published research on M&V 2.0 has focused on applications in the commercial buildings sector (Kupser et al. 2016). In one study, fifteen of sixteen available technologies offering M&V 2.0 capabilities were found to target commercial buildings (Granderson & Fernandes 2017). The accuracy of ten novel M&V 2.0 modelling approaches was evaluated using a data set of 537 commercial buildings and a testing procedure that is particularly well suited for evaluating black-box and proprietary models (Granderson et al. 2016, Granderson & Price 2014). This analysis showed that interval data acquired from AMI offers significant potential for scaling the adoption of M&V using a whole-building approach.



(a) IPMVP approach

(b) Model construction technique

(c) Sectoral applicability breakdown

(d) Input data measurement frequency

(Data source (Granderson & Fernandes 2017)).

Figure 2.6: Graphical illustration of M&V 2.0 applications on the market

A cloud-based platform for estimating energy savings for any ECM in commercial buildings has been developed and applied in a case study (Ke et al. 2017). A limiting factor of this solution is the requirement of the user to input the independent variables, or model features, that are to be used to construct the baseline energy model. This restricts the solution's applicability outside the residential and commercial buildings sectors.

An automated whole-building M&V 2.0 tool was applied to historic data sets from energy efficiency programmes (Granderson et al. 2017). The ASHRAE uncertainty in savings criteria were met in 80% of cases when the automated tool was applied. As with previous research (Yik et al. 2001), the accuracy of the M&V

2.0 approach improved when the buildings were grouped together and treated as a portfolio. Critically, it was concluded from this analysis that uncertainty holds value for evaluating and reducing project and investment risk and it's widespread inclusion in M&V methods may facilitate the effective integration of energy efficiency into regional and national efforts as a tool similar to that of supply-side resources. On a comparably high-level, a means of verifying energy savings from an energy efficiency policy perspective has been developed (Horowitz 2011). In addition, GP modelling has been used to determine energy savings and uncertainty levels in commercial office buildings resulting in significant cost reductions for carrying out M&V (Heo & Zavala 2012).

Most commonly, M&V 2.0 applications apply a data-driven, or black-box, approach to the construction of the baseline energy model. In contrast to this, although remaining in line with the principles of M&V 2.0, a white-box approach utilised the *eQUEST* software to calibrate energy simulation results using the IPMVP Option D (Ke et al. 2013). Also, transfer functions based on internal temperature and consumption models have been applied to perform M&V in an administration building (Díaz et al. 2018). Results highlighted the advantage of maintaining accuracy with minimal monitoring and requiring less data than typical approaches. Advancing the algorithms used for energy modelling is the first step in evolving to M&V 2.0 practices in the industrial sector and it has already been highlighted in Section 2.5 that there is no one single approach that fits best for this application.

2.6.2 Industrial Applications

To date, there have been very few applications of M&V 2.0 principles tailored for the industrial buildings sector. A methodology for using industrial facility-wide billing data to measure energy savings has been developed, although the approach is limited by the information in the data set which is sparse in both the system and time domains (Kelly Kissock & Eger 2008). This methodology implements the Option C as proposed in the IPMVP. A key consideration in this approach is that savings should be greater than 10% of the total site energy consumption. Thus, its application is further limited to ECMs that impact significantly on entire facilities. Projects of this scale can be rare due to the capital that is often required to implement them.

As an alternative to classical statistical approaches, Bayesian statistics have been used for M&V (Carstens et al. 2018*a*, Carstens 2017). This approach has the advantages of resulting in greater transparency, being capable of exactly quantifying uncertainty and its suitability to problems with small data sets. This research was further developed with a guideline for applying these Bayesian techniques within the scope of M&V published (Maritz et al. 2018). Similar guidance is required for classical statistical approaches.

An industrial applications specific methodology for energy savings verification with a case study on a combined heat and power plant has been presented (Rossi & Velázquez 2015). This approach differs from a similar methodology reviewed already (Ke et al. (2017)) in that it pays particular attention to the selection of independent variables. Despite the increased focus in this regards, the process remains manual in nature. It is crucial that this task is automated to fully empower M&V practitioners that do not possess knowledge of the machine learning techniques.

Despite the presence of these solutions for M&V 2.0 in the industrial buildings sector, there is no one solution that enables practitioners to apply advanced analytics on large data sets without possessing knowledge of the underlying algorithms. The application presented in Chapter 6 of this thesis automates the critical steps of the process that require this knowledge including data cleaning, feature selection, model application, model evaluation and deployment.

2.7 Persistence of Energy Savings

2.7.1 Problem Overview

In typical practices, an M&V plan is developed prior to any implementation works taking place. This plan details the ECM and the process that will be followed to verify its performance in the reporting period. A final report is then compiled at the completion of the reporting period. This includes a final project evaluation and savings realised with associated uncertainties. Given that typical reporting periods last for 12-months, there exists a significant oversight in that savings are realised beyond this reporting period. There is a notable opportunity to be realised in increasing the longevity of the M&V process for useful performance verification. To do so, the tasks involved in measuring and verifying energy savings need to be integrated into the ongoing energy management task of M&T. The data-driven M&V 2.0 methods lend themselves to automation, which is vital in tracking energy savings over the lifetime of an ECM.

In the context of individual projects, the lack of performance verification over the lifetime of an ECM presents itself as an opportunity to ensure savings persist. However, on an energy policy level, it causes concern when evaluating the effectiveness of policies. As discussed in Chapter 1 (Section 1.1.3), the foundations of the measures implemented pursuant to the EU's EED are reliant on energy savings persisting beyond their first year of realisation. For example, if savings are counted in Year 1 (2014), they are still expected to be realised in Year 2 (2015). The current M&V protocols do not sufficiently cover the topic of persistence and this highlights a significant knowledge gap in the sector.

There are very few examples of persistence methods in the published literature. Research has been conducted on the persistence of energy efficiency savings and degradation over project lifetime (Skumatz et al. 2009, Hoffman et al. 2015). Despite this, there is no guidance on preventing degradation using FDD in individual projects. A process to continuously measure energy performance using performance indicators is advantageous for persistence, although this is not presented within the scope of M&V (Perroni et al. 2018). A solution tailored to the characteristics of M&V is required to overcome this challenge. The research field of automated fault detection and diagnosis (AFDD) offers the most relatable features from which shared learnings can be identified.

2.7.2 Automated Fault Detection and Diagnosis

AFDD is a process concerned with automating the detection of faults and their causes in physical systems (Katipamula & Brambley 2005*a*,*b*). AFDD represents modern implementations of FDD methods in a manner that is often data centric. There are several published literatures that provide an insight into the broader field of FDD within the context of industrial buildings (Chiang et al. 2001, Russell et al. 2012, Isermann 2006). It is this broader field that founded the methods on which AFDD rely. It is important to note that this section is not intended to comprehensively review the entire research field of AFDD. The objective is to highlight relevant works from which lessons can be learned and techniques adopted for the betterment of the performance verification research field.

There have been a number of review studies that have assessed the current state of the AFDD research field. One such paper presents a comparative analysis of the techniques that are currently employed (Bruton et al. 2014). The key conclusions drawn across the research field are highlighted. These include the view that AFDD tools are no longer seen as standalone solutions. They are being integrated with on-going commissioning tools to enable systems to be returned to their optimal operational performance. In addition, a transition is taking place from early solutions, which were based on single methods of AFDD, to more modern solutions that merge complementary methods to refine the process. Finally, it was concluded that data-driven statistical techniques are growing in application in the field of AFDD. It is findings such as these that must be considered when developing solutions for persistence of savings within the scope of M&V.

It is also pertinent to highlight a number of key applications across the scope of AFDD methods. These methods have been classified initially by Katipamula & Brambley (2005*a*) and this classification matrix has since been expanded on by Bruton et al. (2014). The three high-level classifications are defined as quantitative model, rule and AI-based approaches.

Rule or Physics-based Systems

The effectiveness of a rule-based expert system for performing AFDD in air handling units (AHUs) has been demonstrated (Bruton et al. 2015). Results show it is more effective than early commonly-accepted approaches. Critically, the system developed is capable of prioritising the faults detected based on the impacts on operational costs. Using a more simplistic approach, an easy-to-use FDD tool capable of aiding the identification of faulty equipment based on whole-building HVAC system data was developed (Song et al. 2008). However, a significant drawback of the simplistic approach is the tool's lack of versatility.

Quantitative Model Based

A system that utilises the ASHRAE simplified energy use procedure for fault detection at the whole-building level has been presented. The model was calibrated to measured heating and cooling data to ensure accuracy was maintained. Future weather data was used to forecast the building's energy consumption. Applications in monthly deviations, daily percent deviations and cumulative deviations are discussed (Lee et al. 2007).

Artificial Intelligence Solutions

Bayesian networks have been utilised to detect faults across electrical and mechanical components in AHUs. Bayesian networks are probabilistic graphical models that represent a set of variables and their conditional dependencies via a direct acyclic graph. This approach was deemed to be advantageous in representing and diagnosing complex systems with uncertain, incomplete and even conflicting information (Zhao et al. 2017, 2015). Additionally, ANNs have been used to detect abnormalities in energy consumption, while clustering techniques were used to classify them for diagnostics purposes (Du et al. 2014). The advantage of this approach is the ability to recognise not just known faults, but also unknown faults in the data set. Classification has been further used to detect unexpected energy consumption patterns and perform anomaly detection for specific time windows during the day (Capozzoli et al. 2018). The robustness and flexibility of the methodology has been demonstrated on a data set of whole building electricity loads.

2.8 Conclusions

In this chapter, a comprehensive background to the research field of M&V is presented from the policy mechanisms placing increased responsibility on the process, to the technical shortcomings of current approaches. The role of performance verification in closing the energy efficiency gap was discussed in the context of motivators and barriers to investment in cost-effective ECMs. The threat of rising energy prices, the cost benefits arising from lower consumption and public financing were identified as the primary drivers for investment across all sectors. The nuances between sub-sector drivers were also reviewed in detail. Critically, the main contribution of M&V on the investment level is to the barriers that prevent cost-effective ECMs from being implemented. With risk, uncertainty and hidden costs identified as the most prominent barriers in the energy efficiency sector, the contribution of performance verification in each case highlights the shortcomings of current methods. The theme of these limitations are centred around the retrospective, static methods that employ basic statistical modelling techniques. The concept of energy management was reviewed as it is most often the class to which M&V belongs. Published literature continually concluded that advanced statistical approaches are required for performance verification in optimised EnMSs. The benefits of implementing higher standards to which M&V must adhere to were highlighted in the context of the SEP programme.

Once sufficient background to the wider research field was provided, the technical aspects of implementing M&V were reviewed. As this thesis aims to provide solutions in the industrial buildings sector, a clear distinction was made between the maturity level of research in the residential, commercial and industrial buildings sectors. The established M&V protocols are admired for their robustness and applicability across the broad spectrum of ECMs and industries. However, their limitations in terms guidance given on baseline energy model development, minimising uncertainty and real-time savings verification were discussed. Some opportunities have already been explored in the form of novel approaches. The single biggest opportunity of improving the techniques utilised for baseline energy modelling was reviewed in detail, with data-driven AI methods showing much promise in improving accuracy. Chapter 3 assesses the suitability of these methods to minimise uncertainty with respect to the traditional approaches employed. Following this, Chapter 4 details the development of a prescriptive methodology that provides sufficient guidance for the application of advanced data-driven techniques for energy modelling.

The evolution of practices to a more mature state, known as M&V 2.0, was presented with an evaluation of current solutions concluding that the commercial buildings sector dominates the field. This has led to the industrial buildings sector lagging behind in terms of the methods used to perform M&V. Despite this, the large data sets available as a result of widely available AMI present an opportunity to vastly improve these practices and thus, enabling performance verification to become a powerful resource in ongoing energy management. This topic is revisited in Chapter 5. Finally, the research area of AFDD was leveraged to discover shared learnings that could be applied to develop an energy savings persistence mechanism for M&V. Such a system would offer a solution to many of the challenges that have arisen through energy efficiency policy implementation. In Chapter 6, the findings of all previous Chapters are combined to guide the development of a data science solution to M&V 2.0 in industrial buildings.

Chapter 3

An Assessment of the Suitability of Machine Learning to Minimise Modelling Uncertainty

3.1 Introduction

3.1.1 Overview

The review of published literature presented in Chapter 2 concluded that the development of an accurate baseline energy model is a critical task in individual ECMs, while also having significant repercussions on an energy policy level. The range of methods with which these models can be constructed were evaluated in detail with clear distinctions made between white, grey and black-box approaches. This assessment was carried out beyond the scope of M&V with findings from the broader field of energy modelling included. This proved a fruitful task in identifying the most suitable methods to apply in M&V solutions. In general, white-box models were found to be too complex and computationally expensive for use in performance verification. Although grey-box models were applied successfully in case-study applications, the methods available were deemed to be in their infancy with respect to their more mature counterparts. Finally, it was concluded that black-box models were best suited to exploit the large quantities of energy data available in modern industrial facilities for the purposes of quantifying energy savings.

This chapter investigates the novel use of data-driven machine learning algorithms for M&V of energy savings in industrial buildings. This approach has the important advantage of enabling the extension of the traditional project boundary through the use of more powerful knowledge discovery algorithms. The value of these techniques is evaluated with respect to traditional energy modelling approaches implemented in performance verification. The machine learning regression techniques applied consist of bi-variable and multi-variable OLS linear regression, decision trees, k-nearest neighbours, artificial neural networks and support vector machines. The prediction performances of the models are validated in the context of a biomedical manufacturing facility to find the optimal model parameters. Sensitivity analysis is performed to assess the variance of prediction error as a function of the length of both the baseline and reporting periods.

3.1.2 Background

In 2015, industry accounted for 25.3% of total final energy consumption in the European Union (EU) (Eurostat 2016) and 20.9% in Ireland in 2016 (Sustainable Energy Authority of Ireland 2017). ECMs are being used to reduce this energy consumption that industrial activities are responsible for by optimising the efficiency of energy systems. The term ECM encompasses a wide range of measures and is used to refer to any energy performance improvement project. In recent years, the M&V of energy savings has received increased focus due to measures imposed by energy policy worldwide. Improving efficiency across all elements of energy systems is being utilised as an essential tool to achieve policy targets. Accurate and reliable estimation of energy savings from a wide range of ECMs are needed to cumulatively ensure the effective implementation of the Directive.

To quantify the savings resulting from an ECM, the energy consumption in the reporting period, or post-ECM, must be compared to what the consumption would have been had the ECM not been implemented. This is known as the adjusted baseline. Hence, the post-ECM consumption must be normalised to pre-ECM conditions. In it's most elementary form, this can be identified as a regression problem. Regression analysis consists of quantitative analysis techniques that comprise a set of statistical methods for estimating the relationships amongst variables and subsequently, representing the state of a system. There are two types of regression analyses: simple and multiple regression. Simple regression, often referred to as bi-variable regression, consists of only two variables at any one time. In contrast, multiple regression addresses the relationships between more than two variables at a time. In addition, the relationships in both cases can be either linear or non-linear. Figure 3.1 provides an illustration of the regression problem and the task of fitting a model to the data. This topic is elaborated on in Section 3.3.

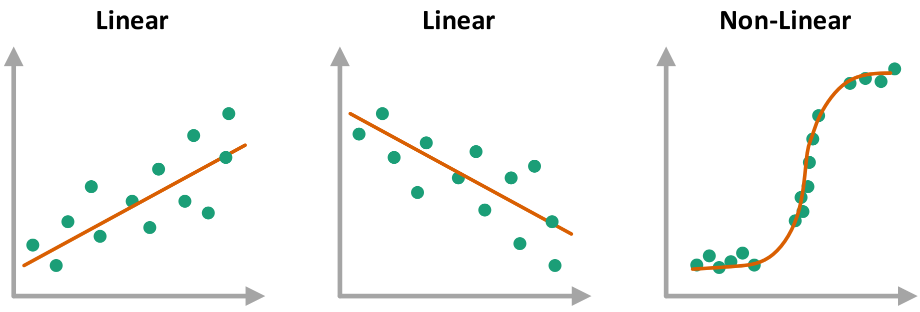


Figure 3.1: Illustration of linear and non-linear regression problems

Engineering or statistical methods are typically employed to construct a baseline energy model capable of performing this normalisation. The accuracy with which the energy savings can be quantified is reliant upon the level of uncertainty that exists. Thus, uncertainty analysis is a necessary step in reliably estimating the energy savings, as an estimation of energy savings alone is insufficient to validate an ECM. A quantifiable measure of uncertainty must also be provided to give an indication of the savings estimation accuracy.

As discussed previously, the three sources of uncertainty in M&V are sampling, modelling and metering. This body of research is concerned with minimising the modelling uncertainty that exists in projects. Estimating the uncertainty in M&V provides a deeper insight into the energy savings and supports the decision making process in developing baseline energy consumption models (Walter et al. 2014). The discrepancies between the methodologies developed by both ASHRAE and EVO have been discussed in Chapter 2 (Section 2.4.3). The approach employed to quantify uncertainty in this research utilises the commonalities between the two most prominent methodologies identified.

Outside air temperature and production levels are often given as common examples of independent variables that can be used to develop baseline models. In machine learning, these independent variables are referred to as features. For many energy systems in industrial buildings, the features that impact energy consumption most significantly can be identified using knowledge of engineering first principles. M&V practitioners are often satisfied to develop baseline models using these most prominent features as they can be employed to achieve reasonable levels of accuracy. However, this results in many relationships between energy consumers not being analysed. Would the inclusion of these features significantly improve the accuracy of energy savings verification? The size of these relationships relative to the more prominent features make them difficult to infer and utilise for the purposes of model development. This chapter investigates the ability of machine learning to utilise these lesser known features with the objective of minimising modelling uncertainty to deliver accuracy and precision. The proposed approach allows the less prominent features to be included in the analysis, thus enabling a wider project boundary to be employed with more efficient data processing techniques.

3.2 Research Questions

A number of previous studies have investigated the suitability of machine learning for modelling baseline energy consumption in end-use residential, commercial and industrial applications. These are reviewed in detail in Chapter 2. As stated previously, there are many commonalities between energy modelling in residential and commercial buildings. However, industrial buildings operate quite differently with complex, multi-faceted energy systems. The deficiency of research on M&V in the industrial sector means critical questions remain unanswered, the following of which are addressed in this chapter:

1. Does a wider boundary of analysis aid the reduction of uncertainty in M&V?
2. Can machine learning be utilised to improve the prediction accuracy for M&V in industrial buildings?
3. How does missing baseline data affect the ability to accurately perform M&V?
4. Can optimal modelling parameters be identified for all use cases?

In addition to these chapter specific research questions, RO1 is the overarching motivator for the work detailed.

3.3 Machine Learning Regression Techniques

3.3.1 Ordinary Least Squares Linear Regression

The OLS approach is the most common model fitting method employed in linear regression (James et al. 2013). It is a type of linear least squares method for estimating the unknown parameters in a linear regression model. OLS chooses the parameters of a linear function of a set of explanatory variables by the principle of least squares, which minimises the sum of the squares of the differences between the measured and predicted values of the dependent variable. The simple linear relationship between two variables, X and Y, is expressed approximately as follows:

$$Y \approx \beta_0 + \beta_1 X \tag{3.1}$$

where, Y is the dependent variable, X is the independent variable (model feature) and β_0 and β_1 are the model coefficients or parameters. In equation 3.1, β_0 and β_1 are two unknown constants that represent the intercept and slope of the straight line that best fits the data respectively. The least squares principle is used to find the values of the parameters which best fit the data. This allows the equation representing the model to become as follows:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \tag{3.2}$$

where, $\hat{y}$ represents a prediction of Y. The *hat* symbol ($\hat{}$) denotes an estimated value. These values are most commonly estimated or fitted using the OLS approach. This is achieved by minimising the sum of the squared residuals, e_i, that are defined in Equation 3.3.

$$e_i = y_i - \hat{y}_i \tag{3.3}$$

The fitting of a simple linear regression model using the OLS algorithm is illustrated in Figure 3.2 using the *mtcars* data set. This data set comprises ten aspects of auto mobile design and performance for thirty-two auto mobiles (Henderson & Velleman 1981). These data have been employed to demonstrate the mechanics of the modelling algorithms utilised in this chapter. The data plotted in Figure 3.2 represents the engine size and associated fuel efficiencies for the cars contained in the data set. The fit is found by averaging the squares of the residuals and minimising this value. Although, this is a very simplistic approach, it is widely used and has been extensively applied within the field of energy modelling as is highlighted in Chapter 2.

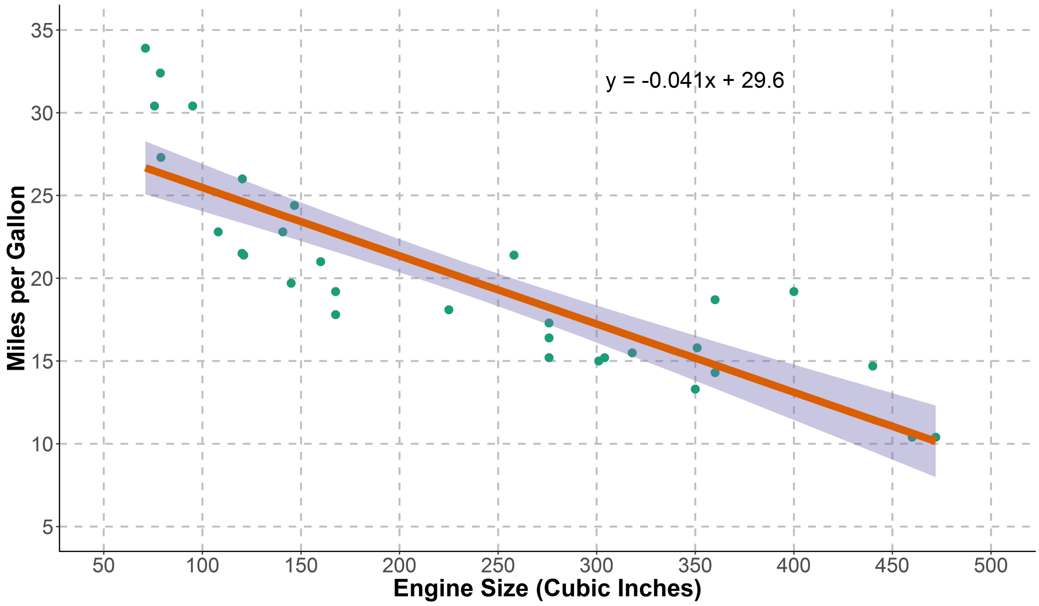


Figure 3.2: Example of fitting simple linear regression model to data set using the OLS algorithm (Shading represents 90% confidence interval)

3.3.2 Decision Tree Regression

Before detailing decision tree supervised learning, the concept of a feature space must be understood. The feature space refers to the n-dimensions in which the variables exists, where n is equal to the number of independent variables or model features. This concept becomes more important when the number of independent variables is large and therefore difficult to visualise. Tree-based methods partition this feature space into a set of rectangles and then fit a simple model in each one (Hastie et al. 2009). They are simple in concept, however, they are powerful when applied in the correct manner and with suitable data characteristics. This process is illustrated in Figure 3.3, which was generated using the *mtcars* data set. In this example, two independent variables are used to model the fuel efficiency (the dependent variable). The independent variables employed are engine power output and the number of engine cylinders. This results in a crude model of the system being fitted for demonstrative purposes. Each internal node represents a test from which the next step is decided. For example, the first node in the tree determines the next step based on the number of cylinders in an engine. Each end note represents an estimated value of the dependent variable.

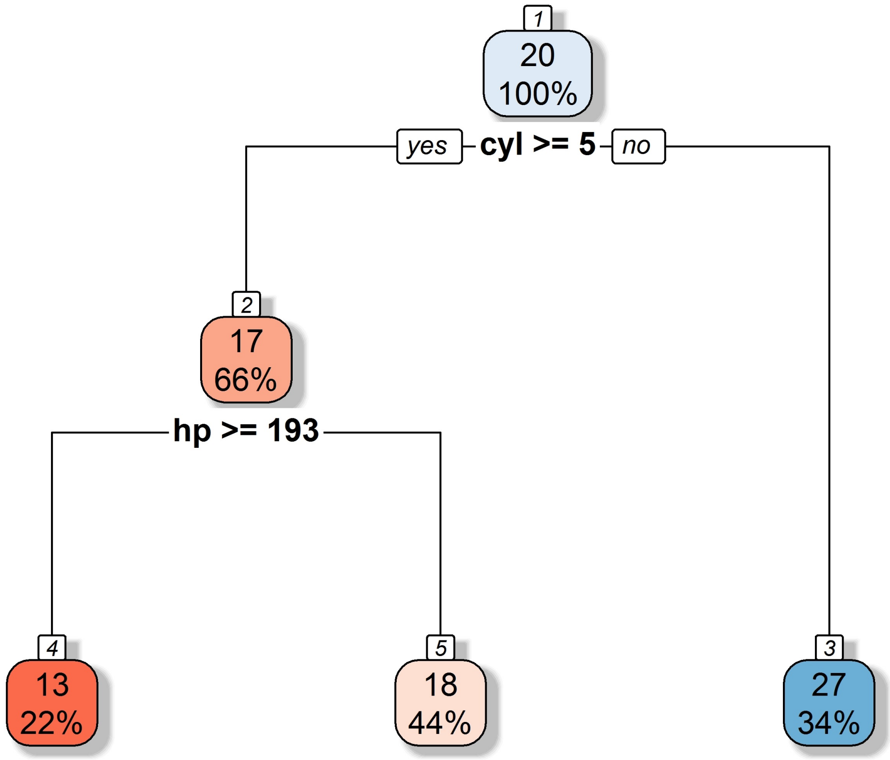


Figure 3.3: Example of fitting a basic decision tree regression model to predict the fuel efficiency of a car stock

There are generally two steps in the construction of a regression tree. Firstly, the feature space is divided into J distinct and non-overlapping regions, $R_1, R_2, ..., R_J$. As with the OLS regression approach, the divisions in the feature space are chosen with the goal of minimising the residual sum of squares (RSS) defined in Equation 3.4. Following this, the same prediction is made for every observation that falls into the region R_J. The prediction made is the mean of the dependent variable values for the training observations in R_J.

$$RSS = \sum_{j=1}^{J} \sum_{i \in R_J} (y_i - \hat{y}_{R_J})^2 \tag{3.4}$$

where, $\hat{y}_{R_J}$ is the mean response for the training observations within the J^{th} box. As it is not computationally feasible to consider every possible partition of the feature space into J boxes, a recursive greedy algorithm is utilised for creating the splits. The process is termed recursive because each sub-population created by each split may in turn be split an indefinite number of times until the process is terminated by a stopping criteria. This stopping criteria is dependent on the specific algorithm employed. These algorithms include the classification and regression tree (CART), patient rule induction method (PRIM) and C4.5 approaches. It is common to stop splitting when some minimum node size is reached or when a split does not improve performance by more than a given threshold. Tree pruning is then carried out using such approaches as the cost-complexity pruning method.

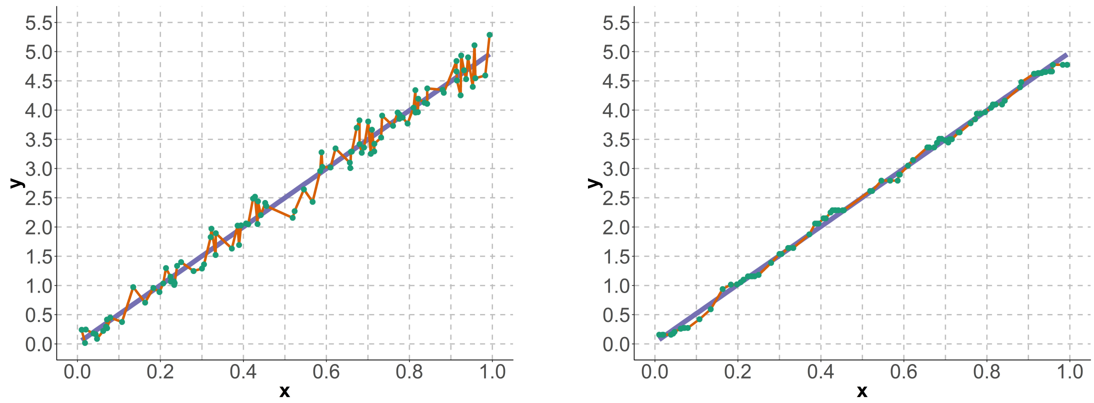
3.3.3 k-Nearest Neighbours Regression

K-NN regression is one of the simplest and best-known non-parametric supervised learning methods (James et al. 2013). They are an example of an instance-based learner which are in contrast to the methods discussed already in which some fixed set of parameters are learned. Instance-based learners compare new problem instances with instances seen in training data that are stored in memory. They do not perform explicit generalisation. Following the specification of a value for K and given a prediction point x_0, k-NN regression initially identifies the K training observations that are closest to x_0, represented by N_0. It then estimates $f(\hat{x_0})$ (Equation 3.5) using the average of all the training responses in N_0.

$$f(\hat{x_0}) = \frac{1}{K} \sum_{x_i \in N_0} y_i \tag{3.5}$$

As illustrated in Figure 3.4, a higher value of K results in a smoother fit. The optimal value of K is found by balancing the trade-off between bias and variance. This can be optimised using methods such as cross-validation which were reviewed in Chapter 2 (Section 2.4.3). K-NN is also capable of modelling non-linear relationships (James et al. 2013).

In cases where more than one independent variable exists, the data set must be standardised. This is required as variables that are measured using different units and scales can skew the calculation of distance to nearest neighbours. Finally, the method employed to calculate the distance between data points can also impact on performance. The Euclidean distance (Equation 3.6) is the most widely used distance function and it has been found to perform well on both categorical and numerical data sets (Hu et al. 2016). It is also important to note that k-NN can be computationally expensive to train, which represents a significant drawback



(a) $K = 1$ (b) $K = 9$

Figure 3.4: Illustration of the impact differing values of K have on model fit and smoothness. In both figures, $f(\hat{x_0})$ (green) is plotted using a data set containing only one independent variable, x. y and x are linearly correlated with the purple line showing their directly proportional relationship.

to an otherwise straightforward approach.

$$dist(A, B) = \sqrt{\frac{\sum_{i=1}^{m}(x_i - y_i)^2}{m}} \tag{3.6}$$

where, A and B represent feature vectors and m is the dimensionality of the feature space.

3.3.4 Artificial Neural Network Regression

The term neural network, or ANN, encompasses a large class of models and learning methods. The single hidden layer back-propagation network, or single layer perceptron, is the most widely used neural network (Figure 3.5) (Hastie et al. 2009). An ANN is an information processing paradigm that has been inspired by the way the human nervous system works to process information. The term back-propagation refers to the method used to train the model. An ANN comprises a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems (Stergiou & Siganos 2013). In regression applications, there is typically only one output unit (or node) in the network, however, these networks can handle multiple quantitative responses.

There are many different types of neural networks ranging from feed-forward approaches to the more complex recurrent neural networks which propagate data both forwards and backwards. For the purposes of this chapter, the feed-forward

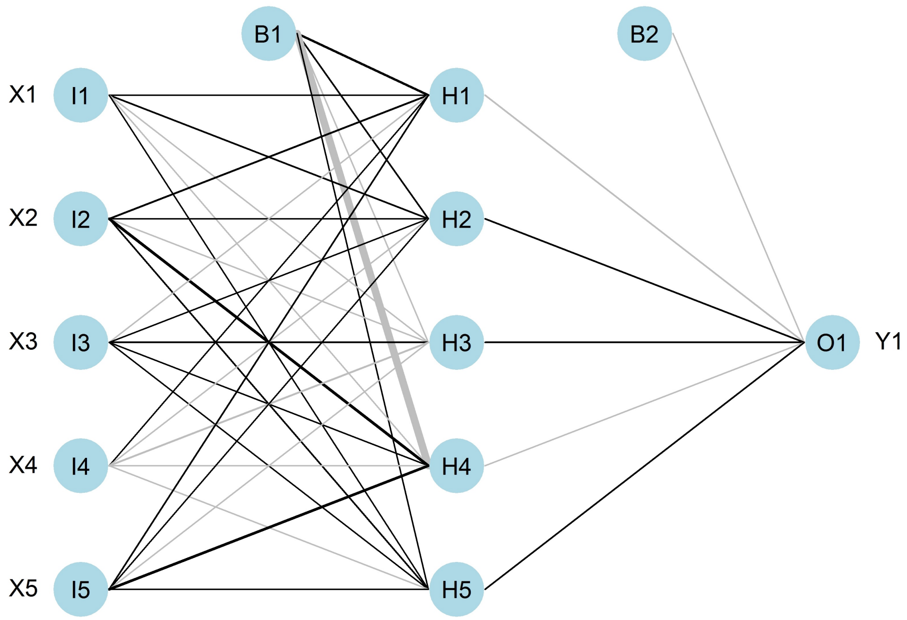


Figure 3.5: Sample structure of feed-forward artificial neural network with one hidden layer containing 5 neurons

approach is focused on as this is the method employed for the analysis detailed. This approach was chosen as it is capable of representing complex functions with relative ease and the reduced computational effort and memory capacity required to train a model (Malinowski et al. 1995).

Using the back-propagation approach to fit an ANN model, random weights are assigned to all parameters at the beginning of the process. Each neuron multiplies this initial weight by the values input into it and the sum of these results is output after being adjusted for the bias of that neuron. An activation function is used to determine whether or not the output is passed onto the next stage in the network. The activation function is usually chosen to be the *sigmoid* $\sigma(v) = 1/(1+exp^{-v})$. Gaussian radial basis functions are also used for the activation function, which produce a radial basis function network. When selecting the unknown parameters or weights, the sum of squared errors is again used as a goodness-of-fit measure. The gradient descent algorithm is the general approach employed to minimise this error by learning the weights and biases throughout the process. A convergence criteria is used to determine when to terminate the fitting process. A learning rate is a hyper-parameter that is used to control the level of adjustment of the weights during the fitting process. The lower the value, the slower the process is.

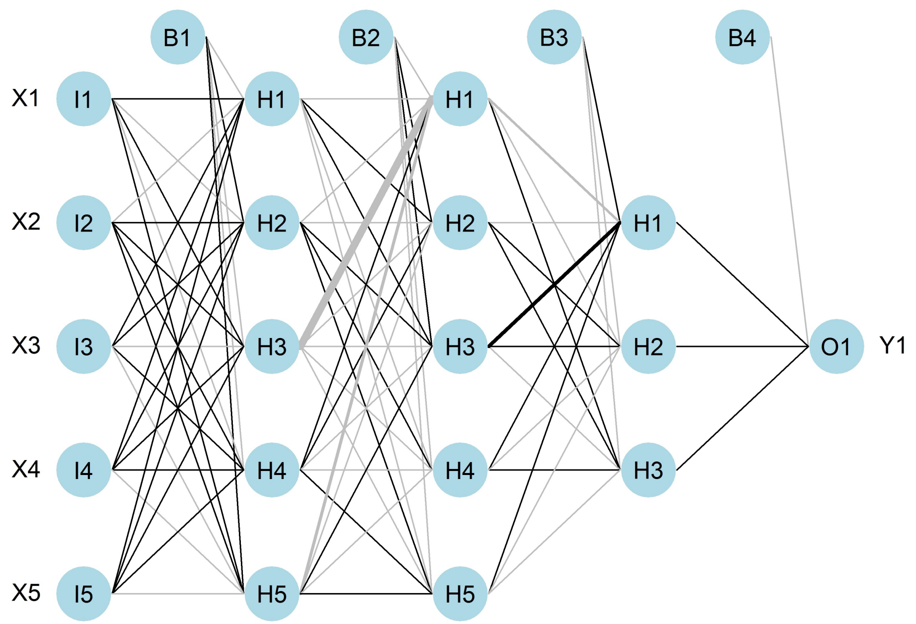


Figure 3.6: Sample structure of feed-forward artificial neural network with three hidden layers each containing a pre-defined number of neurons

There are also a number of additional hyper-parameters that need to be optimised. These include the number of neurons in each hidden layer, the number of hidden layers, learning rate, decay rate, step sizes and activation function type to name a few. These are dependent on the approach implemented and are usually trained using methods such as 10-fold cross-validation.

The back-propagation approach offers the advantages of being simple and possessing a local nature. Each hidden unit passes and receives information only to and from units it shares a connection. However, it can be very slow to train and hence, it is not usually the method of choice (Hastie et al. 2009). Other training methods include genetic and trial and error approaches.

3.3.5 Support Vector Machine Regression

SVM is a supervised machine learning algorithm which can be used for both classification and regression problems. The SVM regression algorithm uses the same principles as that of it's classification counterpart with only a few minor differences. In classification problems, SVM simply finds a hyper-plane that differentiates between two or more classes, with a hyper-plane being defined as a subspace whose dimension is one less than that of its ambient space. SVMs are an example of kernel-based learning methods which belong to the family of instance-based learners.

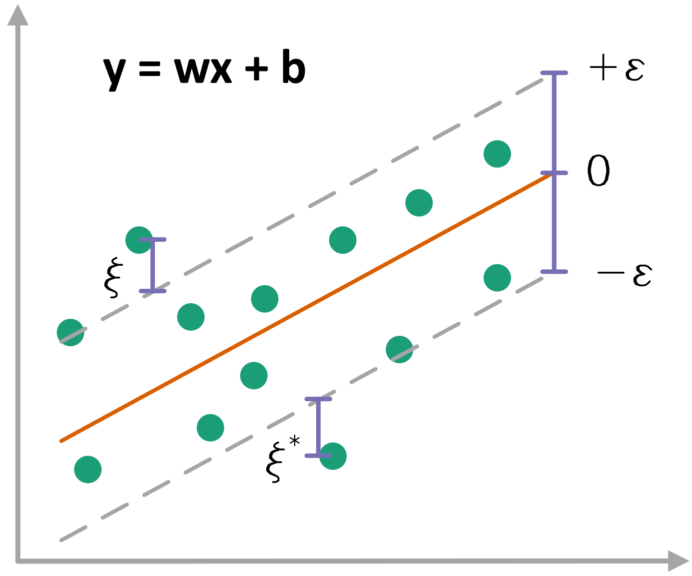


Figure 3.7: Illustration of support vector machine optimisation problem

In contrast to OLS regression, decision trees and ANN which minimise the training error, SVMs attempt to minimise the generalisation error bound so as to achieve generalised performance (Basak et al. 2007). The objective is to find a function $f(x)$ that has at most ϵ deviation from the actually obtained targets, y_i, for all the training data, while remaining as flat as possible concurrently (Vapnik 2000). Equation 3.8 represents the formulation of the algorithm as a convex optimisation problem (Smola, A and Schölkopf 2004). This is formed on the basis that the training data is given $\{(x_1, y_1), ..., (x_\ell, y_\ell)\} \subset X \times \mathbb{R}$, where X denotes the space of input patterns and the $f(x)$ takes the following form:

$$f(x) = (w, x) + b \text{ with } w \in X, b \in \mathbb{R} \tag{3.7}$$

$$\begin{aligned} &\text{mimimise} && \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{\ell}(\xi_i + \xi_i^*) \\ &\text{subject to} && \begin{cases} y_i - (w, x_i) - b \leq \epsilon + \xi_i \\ (w, x_i) - b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned} \tag{3.8}$$

where, ϵ is the threshold, ξ_i and ξ_i^* are slack variables that cope with otherwise infeasible constraints of the optimisation problem and C is a constant that determines the trade-off between the flatness of $f(x)$ and the amount up to which deviations larger that ϵ are tolerated.

Both linear and non-linear models can be developed with the kernel function differing between both applications. Examples of such kernels include linear, Gaussian, and polynomial functions (Equations 3.9 to 3.11). In non-linear applications, the kernel functions transform the data into a higher dimensional feature space to make it possible to perform linear separation.

$$\text{Linear kernel: } G(x_j, x_k) = x_j \cdot x_k \tag{3.9}$$

$$\text{Gaussian kernel: } G(x_j, x_k) = \exp(-\left\|x_j - x_k\right\|^2) \tag{3.10}$$

$$\text{Polynomial kernel: } G(x_j, x_k) = (1 + x_j \cdot x_k)^q, \text{where } q \text{ is in the set } \{2, 3, ...\} \tag{3.11}$$

A significant benefit in using the SVM algorithm is that optimality is guaranteed. The nature of convex optimisation ensures a global minimum is found, as opposed to a local minimum. Additionally, feature mapping is implicitly carried out. The SVMs detailed in this thesis were all developed using a linear kernel. This was chosen to minimise computational time and complexity.

3.4 Methodology

As outlined in Section 3.2, the objective of this analysis is to assess the suitability of machine learning for improving M&V in industrial buildings. As was concluded by Zhao and Magoulès, each model must be compared under the same circumstances for a complete analysis (Zhao & Magoulès 2012). Hence, a biomedical manufacturing facility was chosen as a test-bed for this proposed approach. The site has an approximate footprint of 4.2 acres and has over 1000 employees. It comprises processes such as casting, milling, grinding and packaging. As a result, it utilises a significant quantity of energy in both machines directly, as well as in the preparation and conditioning of clean-rooms to enable medical device manufacture. The site was deemed applicable as no ECM was performed during the period of analysis. This allowed for complete evaluation of the model prediction performance; something that would not be possible had the consumption changed as a result of an ECM. The building characteristics (e.g. envelope, materials, openings) did not change during the period of analysis, hence, these factors were not considered in the baseline energy model.

The compressed air electricity consumption at the site was selected to be modelled for this investigation. ECMs are very commonly performed on compressed air systems as they can achieve savings of 20-50% on average (Saidur et al. 2010). In this case study, the total electricity consumption of the compressors was metered, but there was no flow meter on the compressed air main header. To carry out M&V on this system, there are two clear options. One solution is to install a meter to measure compressed air flow. This would be useful for quantifying savings on the generation side of the system, although it would increase costs and delay the project as baseline data would need to be gathered. The alternative solution is to model the compressed air electricity consumption based on it's relationships with other energy consumers within the facility. This approach requires the construction of a model of baseline consumption; thus, it was deemed an ideal case study for conducting the analysis.

There were 24 months of data available for this study. The data set was split to hold out 12 months of data for training the models and the remaining 12 months stored separately to be used as a testing data set for model evaluation on unseen data. These would be representative of baseline (pre-ECM) and reporting (post-ECM) periods in a practical application. It is important to note that although an 80/20 split of training to test data is more common in machine learning, a 12-month testing period is representative of real world M&V applications. This approach has been developed and applied previously by Granderson et al. and is effective in simulating the conditions necessary for comparison of model performance (Granderson & Price 2014). Table 3.1 contains a summary of the data available to be used in the modelling process. The input variables employed contain a mix of both building and process related energy consumers. These are classified in Table 3.1 and ensures that both building and process related energy consumption is accounted for in the models, thus providing an accurate representation of the system operation in the baseline period. All analysis was carried out using the open source statistical programming language R.

Table 3.1: Summary of variables included in the available dataset.

Variable	Description	Average Value	Type	Service
Compressed Air	Total electricity consumption of four air compressors.	363.3 kW	Dependent	Process
Chilled Water	Total electricity consumption of chilled water generation system.	142.7 kW	Predictor	All
Heating	Electrical heating load.	36.4 kW	Predictor	All
Cooling Tower Water Pumps	Electricity consumption of cooling tower water pumps.	5.9 kW	Predictor	All
Dust Extraction	Total electricity consumption of dust extraction system.	87.6 kW	Predictor	Process
Grid Electricity	Quantity of electricity imported from the national grid. On-site generation services the remainder of the load.	1747 kW	Predictor	All
Production HVAC	Electricity consumption of HVAC servicing production floor area.	82.21 kW	Predictor	Process
Non-Production HVAC	Electricity consumption of HVAC servicing all non-production areas.	28.3 kW	Predictor	Building
Production	Production equipment electricity consumption.	1355 kW	Predictor	Process
Outside Air Temperature	Outside air temperature measured in degrees Celsius.	°15C	Predictor	Building
Operation	Status of operation in the facility (1 = In-production, 0 = On-standby).	-	Predictor	Process

3.4.1 Algorithms

Five prominent machine learning algorithms were selected to solve this problem; multi-variable linear regression, decision tree regression, k-nearest neighbours, artificial neural networks and support vector machines. There are a wide range of algorithms that could be applied for this type of analysis; however, these five were selected based on previous success in the field in the published literature reviewed in Chapter 2 (Section 2.5). For comparative purposes, it was decided that an OLS regression model constructed using outside air temperature and production electricity was a reasonable assumption of a typical approach taken by M&V practitioners.

In machine learning, the term hyper-parameter is used to distinguish from standard model parameters. Standard model parameters are learned in the model training process. However, hyper-parameters cannot be directly learned from the regular training process. These parameters convey properties of the model such as its complexity and the speed of learning. The optimised value of each hyper-parameter was found by performing a grid search on possible values and using 10-fold cross-validation on the training data to determine the best performing model. 10-fold cross-validation was deemed an appropriate means to estimate prediction error based on published research (Zhang & Yang 2015, Kohavi 1995). It also prevents the modelling algorithms from over-fitting to the training data. In addition to this, the ANN weight decay hyper-parameter is used to prevent over-fitting.

The optimised hyper-parameter values are specific to each individual application and thus, allow the methodology to be adaptable and customisable to the properties of any given data set. Descriptions of the machine learning algorithms applied, the hyper-parameters associated with each and the notation used throughout this paper can be found in Table 3.2.

Table 3.2: Description of machine learning algorithms employed in the analysis

Algorithm	Description	No. of Features	Hyper-parameters	Grid Search	Notation
Two-variable Linear Regression	An ordinary least squares approach assumed to be representative of typical M&V practice. Production electricity consumption and outside air temperature are the features employed.	2	Intercept	True/False	Bi-Lin
Multi-variable Linear Regression	A more detailed ordinary least squares model constructed using 9 additional features from the available data set.	11	Intercept	True/False	Multi-Lin
Decision Tree Regression	Models in the form of a tree structure with decision nodes. The topmost node in a tree represents the best predictor.	11	Maximum tree depth	d_{max} = 1:10	Tree
k-Nearest Neighbours	Non-parametric model where the input consists of the k closest training examples in the feature space. The output is the average of the values of its k-nearest neighbours.	11	Maximum no. of neighbours Distance Kernel	k_{max} = 1:10 d = 1:5 $kernel$ = rectangular, triangular	k-NN
Artificial Neural Networks	Non-linear statistical model. It is a two-stage regression model typically represented by a network diagram. A single hidden layer feed-forward neural network was developed in each instance.	11	No. of hidden units Maximum no. of iterations Threshold Weight decay	$size$ = 1:11 it_{max} = 500,000 t = 0.01 d = (0.5,0.1,0.01,0.001)	N-net
Support Vector Machines	Non-parametric technique reliant on kernel functions. Examples are represented as points in space with a clear gap separating mapping categories.	11	Kernel Cost	$kernel$ = linear c = (0.25,0.5,1,10)	SVM

3.4.2 Performance Metrics

The models trained using the baseline period data were applied to the testing data to evaluate prediction performance. The CV(RMSE) is a measure of the variability between the actual and predicted values. It is calculated by dividing the root mean square error by the average energy consumption (American Society of Heating Refrigerating and Air-Conditioning Engineers 2014). The CV(RMSE) is a metric used to quantify modelling error in both ASHRAE Guideline 14 and IPMVP. The equation for the metric is provided in Equation 3.12, where y_i is the actual value, $\hat{y}_i$ is the predicted value, $\bar{y}$ is the average of the actual value, n is the total number of predictions in the period of analysis, and k is the number of model features. In the ASHRAE Great Energy Predictor Shoot-out II, CV(RMSE) was the primary metric employed to determine overall model ranking.

$$CV(RMSE) = \frac{1}{\bar{y}} * \sqrt{\frac{\sum_i^n (y_i - \hat{y}_i)^2}{n-k-1}} * 100 \tag{3.12}$$

In the same study, NMBE was the secondary metric used to support the evaluation process (Haberl & Thamilseran 1996). The mean bias error is an indication of overall bias in a regression model and is calculated using the formula in Equation 3.13. It quantifies the tendency of a model to over or underestimate across a series of values. This metric is independent of time-scale so care must be taken as overall positive bias error can cancel out negative bias. In contrast, the CV(RMSE) does not suffer from this problem.

$$NMBE = \frac{1}{n-1} * \frac{\sum_i^n (y_i - \hat{y}_i)}{\bar{y}} * 100 \tag{3.13}$$

The median of the absolute relative error (med(absRTE)) is a useful metric to understand the typical error in the prediction of total energy consumption over the testing period. The metric is similar to the mean absolute percent error, but uses the median to overcome the sensitivity of the mean to extreme values. Equation 3.14 contains the formula for calculating the med(absRTE).

$$med(absRTE) = median(\frac{abs(y_i - \hat{y}_i)}{y_i}) \tag{3.14}$$

3.4.3 Model Uncertainty

The procedure for calculating the uncertainty introduced by the baseline model is explicitly defined by both IPMVP and ASHRAE Guideline 14. In both cases, the CV(RMSE) is used, along with other measures, to compute the uncertainty associated with the model. The formulae used to calculate this uncertainty varies between IPVMP (*Statistics and Uncertainty for IPMVP*) and ASHRAE Guideline 14. However, the CV(RMSE) is common to both cases, with all other equation parameters being independent of the model constructed. These parameters include t-statistics, sample size and the number of independent variables. As CV(RMSE) is the only parameter effected by the model performance, it is important to focus on minimising it to achieve the objectives stated in Section 3.2.

In a practical application using IPMVP approaches, the CV(RMSE) is calculated by applying the baseline model to the pre-ECM dataset (i.e. applying the model to the data used to train it). This approach is very susceptible to over-fitting the model to the training data. To overcome this issue, a building in which no ECM has been implemented was chosen as a test site. A truer measure of performance can be found by applying the baseline model to the testing data. This model validation procedure allows for direct comparison of the adjusted baseline, calculated by the model, and the measured post-ECM consumption. This is an implementation of the approach previously developed by Granderson et al. (Granderson & Price 2014).

3.5 Results and Discussion

3.5.1 Potential of Additional Model Features

A notable characteristic of the analysis is the use of additional model features that would otherwise be overlooked. Section 3.1.2 describes the relevance of these features and the typical approach taken by M&V practitioners. It was deemed that a typical approach would use production electricity consumption and outside air temperature as the predictors (model features). This assumption is based on correlation analysis and engineering first principles, which are commonly employed techniques in M&V.

Analysis was carried out to assess the value in employing 9 additional features in the model construction process. Hence, the second approach employed all 11 features that were available in the facilities data set. Baseline energy models were developed using an OLS regression algorithm for both the traditional (2 model features) and proposed (11 model features) approaches. 12 months of training data and a selection of measurement frequencies were used in the analysis. The performance of each model was evaluated using 12 months of unseen testing data.

Figure 3.8 illustrates the results of this analysis. For daily, hourly and quarter-hourly measurement frequencies, it was found that the models developed using all 11 model features outperform those constructed using the more traditional approach. This is not the case when less granular weekly and monthly data is employed. The more straightforward approach performs best in these instances. It is the hypothesis that the more complex model is more reliant on the knowledge contained within the additional variables, which is diminished at lower measurement frequencies. The three best performing models across the spectrum of temporal granularities are those developed using all 11 model features. These would ordinarily not be employed for this analysis using current methodologies. The use of these additional features expands the boundary of analysis; thus, offering a novel and more accurate means of achieving the objectives of M&V. The best performing model overall uses all 11 features and a 15-minute measurement frequency. In comparison to the most accurate traditional model, CV(RMSE) and NMBE are reduced by 15.9% and 75.6% respectively.

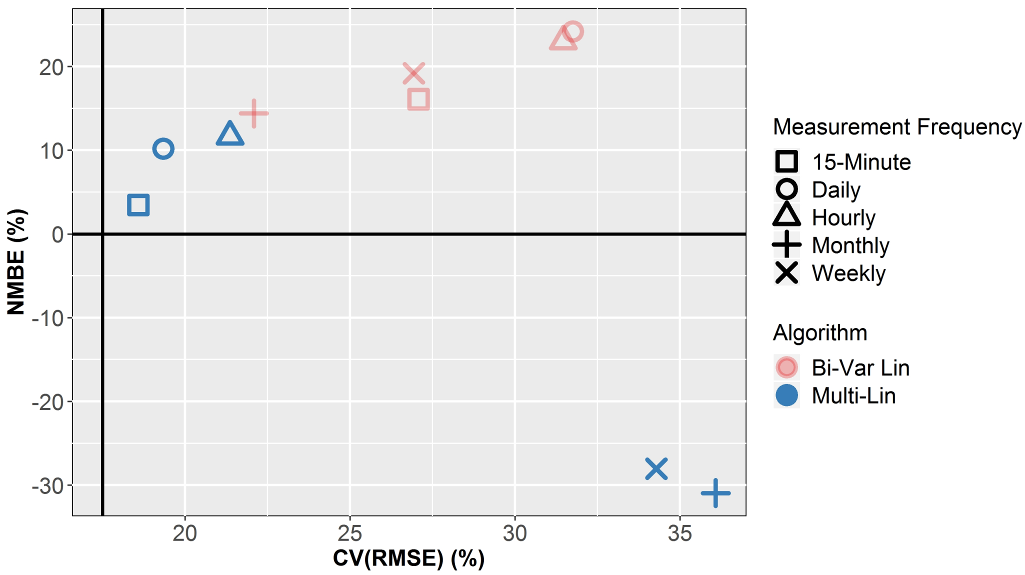
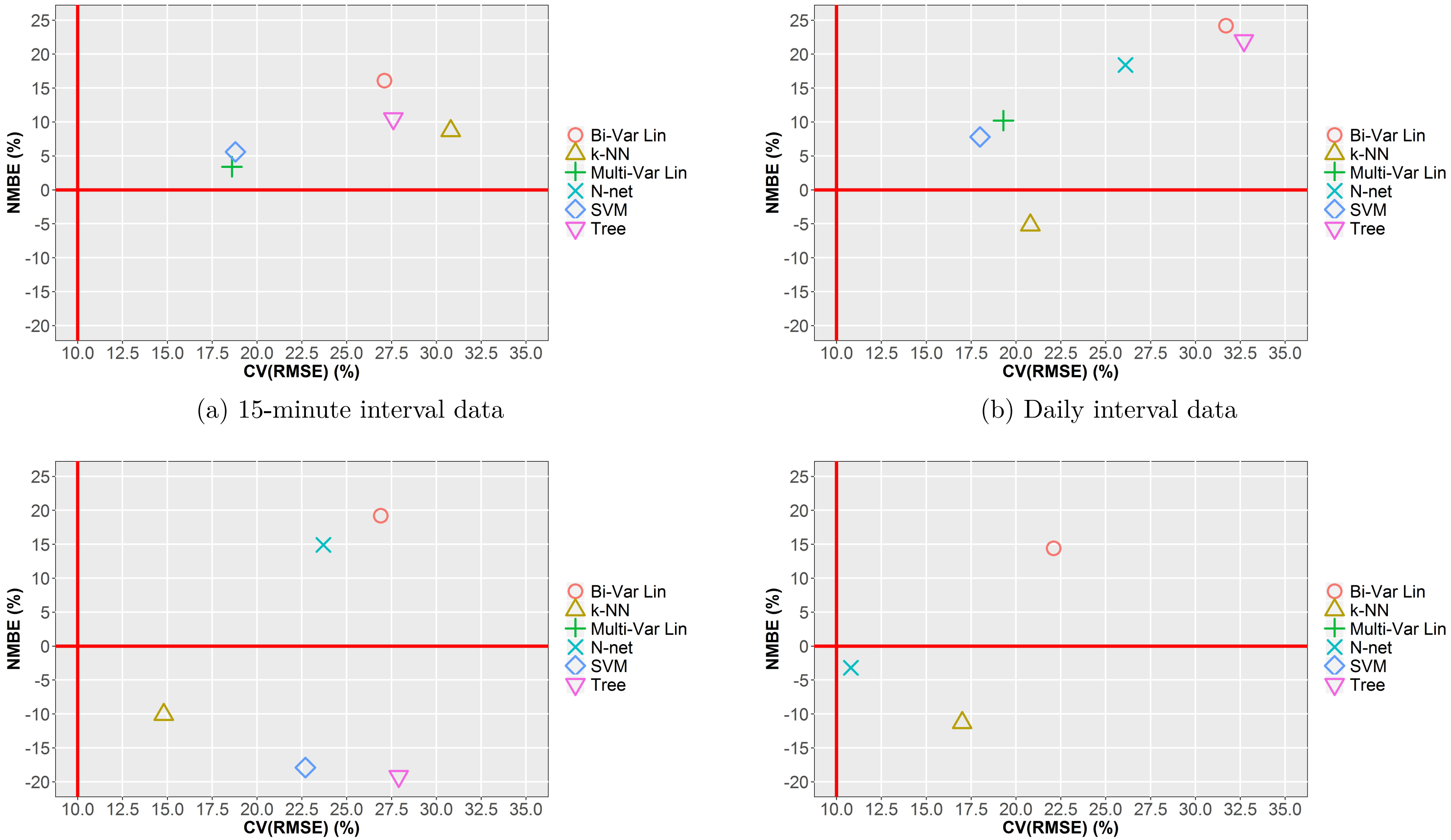


Figure 3.8: Assessing the value of additional model features for different temporal granularities

3.5.2 Harnessing the Power of Additional Features

The findings in Section 3.5.1 identified the potential benefits of expanding the scope of analysis beyond the currently employed techniques. Further analysis investigated the ability of different algorithms to improve the accuracy in estimating the adjusted baseline, beyond that of the OLS model constructed using the 11 features. The training period was held constant at 12 months with monthly, weekly, daily and quarter-hourly measurement frequencies reviewed. The traditional two-variable approach used previously was again employed for comparative purposes, with all other modelling algorithms utilising all 11 features.

Figure 3.9 contains a graphical representation of the model performance for each set of project parameters with comprehensive performance metrics included in Table 3.3. The use of a higher measurement frequency does not always improve the performance of each model; however, it does reduce the spread of error between all 6 models. Monthly data provided the most accurate model across the analysis, a feed-forward ANN with -3.24% NMBE and 10.8% CV(RMSE). In contrast, the OLS, decision tree and SVM models, constructed using all 11 features, performed with unacceptable levels of accuracy at this measurement frequency. The k-NN model constructed using weekly interval data predicted the adjusted baseline with the next best accuracy. The performance of the ANN and k-NN models deteriorates significantly as measurement frequency increases. In contrast to this, the performance of the multi-variable OLS model becomes significantly more accurate as the measurement frequency increases. SVMs become more prominent as the measurement frequency increases also. Decision trees perform poorly across all conditions. The algorithm was unable to sufficiently construct a model at lower measurement frequencies, possibly as it is too simplistic a means of modelling this complex system. The performance of the ANNs are erratic, with the greatest accuracy achieved at lower measurement frequencies.



(a) 15-minute interval data

(b) Daily interval data

(c) Weekly interval data

(d) Monthly interval data

Figure 3.9: The performance of each algorithm using 12 months training data and 12 months testing data

Table 3.3: Performance of each model developed using 12 months training data and evaluated using 12 months testing data with varying measurement frequency

	Two-variable OLS	Multi-variable OLS	Decision Tree	k-NN	ANN	SVM
Monthly						
NMBE	14.39	-30.95	-	-11.32	-3.24	-33.26
CV(RMSE)	22.08	36.08	-	17.00	10.80	41.46
med(absRTE)	8.41	27.7	-	14.2	6.89	43.3
Weekly						
NMBE	19.19	-28.03	-19.23	-10.13	14.93	-17.86
CV(RMSE)	26.93	34.28	27.93	14.82	23.67	22.72
med(absRTE)	21.58	28.36	18.89	8.06	12.74	17.58
Daily						
NMBE	24.20	10.17	22.01	-5.23	18.40	7.85
CV(RMSE)	31.75	19.34	32.73	20.81	26.15	18.04
med(absRTE)	27.17	13.65	29.75	12.45	21.95	13.24
Quarter-Hourly						
NMBE	16.10	3.45	10.50	8.70	17.43	5.62
CV(RMSE)	27.07	18.58	27.63	30.82	36.33	18.80
med(absRTE)	19.64	12.32	22.06	20.97	23.09	13.09

A comprehensive and complete analysis of all approaches under the same set of operating conditions enables an accurate comparative review to be carried out. There is no clear most appropriate modelling algorithm across all four measurement frequencies. Therefore a conclusion cannot be drawn on the most accurate machine learning algorithm for modelling baseline energy in M&V, although there is a clear best performing model for each measurement frequency. Despite this, only one model is required for the purposes of any M&V project. The optimal model is the ANN that uses 11 model features and monthly interval data. This analysis has highlighted the need to conduct this type of exhaustive analysis in each case as individual project requirements and characteristics will influence model performance.

3.5.3 Sensitivity Analysis: Training and Testing Period Length

In Section 3.5.1 and 3.5.2, a 50:50 ratio of training to testing data was applied. Although this is uncommon in most machine learning applications, it is representative of typical M&V cases in which only 12 months of baseline data are available and the adjusted baseline must be predicted for a 12-month reporting period. The algorithms and approach proposed thus far have been proven to be suitable for improving the accuracy of M&V. However, the sensitivity of this accuracy to the quantity of training data available offers further scope to evolve M&V practices in industrial applications. Data availability is an ever present constraint in many M&V projects. To adhere to the IPMVP, backfilling of missing data cannot be carried out in the baseline period. A lack of data in the baseline period is often the single biggest hindrance to completing accurate M&V. Additional metering infrastructure is usually installed to overcome this issue, but this increases project costs and delays implementation as baseline data must be gathered.

To simulate the conditions of missing data, the best performing algorithms from Section 3.5.2 were applied to construct models based on limited training data. The best performing model using the proposed approach employed the ANN algorithm and a monthly measurement frequency. The algorithm hyper-parameters were optimised using 10-fold cross validation, resulting in a hidden layer with 11 neurons and a weight decay of 0.5. For comparative purposes, the best performing model developed using the traditional approach was also brought forward for analysis. This also used a monthly measurement frequency, while employing the OLS linear regression algorithm and just two predictor variables (outside air temperature (OAT) and production electricity). Both models were evaluated using 3, 6, 9 and 12 months testing data. For reduced training periods, the most recent period of data was considered in each case. The practicalities and requirements of M&V limit its accuracy and hence, these unconventional training to testing ratios need to be investigated to fully understand the limitations of the approach.

The sensitivity of the two-variable model is illustrated in Figure 3.10. It is clear that the length of training period directly improves prediction accuracy in every case. This analysis shows the dependency of the traditional approach on the availability of baseline data to train the model. This restricts the potential applications in M&V. The results of the sensitivity analysis conducted on the ANN are included in Figure 3.11 and offer an intriguing insight into the potential of the proposed approach. The models constructed using shorter training periods are capable of performing adequately with respect to those constructed using longer training periods. It is very common in M&V that models are required to predict the adjusted baseline for a 12 month period. This is akin to that of a 12 month testing period in this analysis. For these conditions, the model constructed using 6 months training data performs with a CV(RMSE) of 10.8%, while the model constructed using 11 months training data results in a CV(RMSE) of 10.7%. The prediction accuracy achieved, using almost half

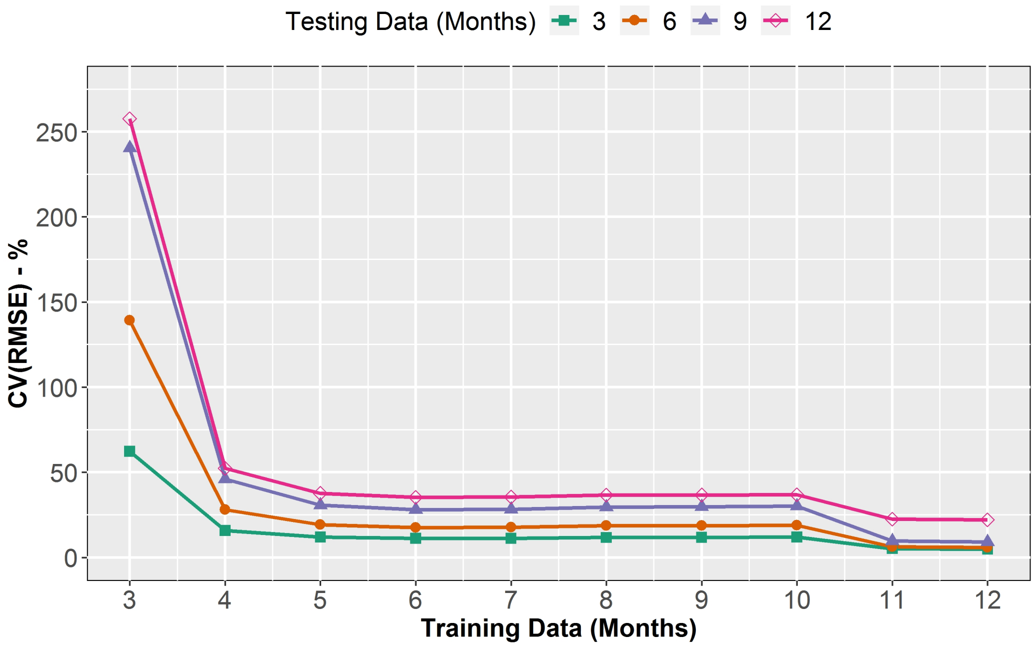


Figure 3.10: Sensitivity analysis of two-variable ordinary least squares linear regression model performance to quantity of training and testing data

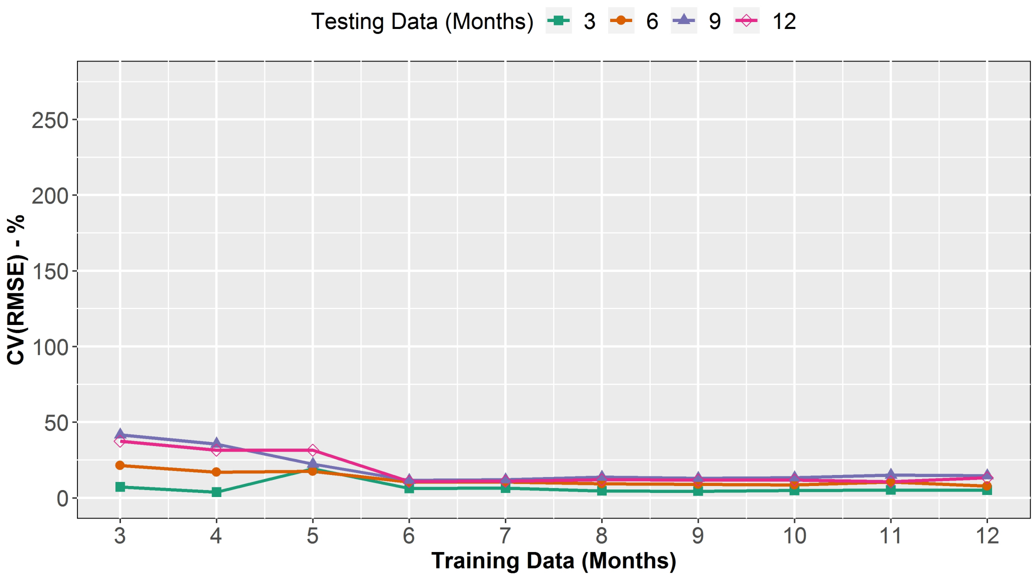


Figure 3.11: Sensitivity analysis of artificial neural network model performance to quantity of training and testing data

the quantity of training data, highlights the potential of the proposed machine learning-based approach to be applicable to projects with limited data available. This pattern in performance is common across the 6, 9 and 12 month testing data sets. Across all testing data sets, it is clear that acceptable performance, relative to that of a 12 month training period, can be achieved using 6 months of training data or more. The results for the testing set containing 3 months of data show promising results, with a CV(RMSE) of just 3.7% achieved using 4 months of training data. This pushes the limitations of traditional techniques to a wider spectrum of applications, while also improving prediction performance and hence, minimising uncertainty.

3.6 Conclusions

Machine learning techniques were found to be an excellent means of minimising uncertainty in industrial applications of M&V. The suitability of five different machine learning techniques were examined with respect to an assumed typical approach. This analysis was carried out in the context of a biomedical manufacturing facility, as it has already been proven that operating conditions must be kept constant for each technique to enable a complete investigation (Zhao & Magoulès 2012). The results identify the potential performance improvements that are achievable by extending the boundary of analysis and incorporating additional independent variables into the model construction process. Machine learning was used as a tool to extract the knowledge contained within the data for these variables and construct models of the baseline energy consumption with varying degrees of success. The use of data-driven modelling enables a dynamic and flexible approach be taken to a wide range of projects.

Section 3.5.1 highlights the accuracy improvements that can be achieved by employing additional features in the analysis, i.e. extending the typical project boundary. The prediction accuracy was improved when the measurement frequency was daily or higher. The CV(RMSE) and NMBE were reduced by 15.9% and 75.6% respectively, when the best performing model constructed using all 11 features is compared to that of the more traditional, two-variable approach. This initial analysis showed promise and hence, the same methodology was applied with four other modelling algorithms in an attempt to further improve the prediction accuracy. An exhaustive methodology was applied to construct the models for varying measurement frequency. This was necessary to identify the optimal modelling algorithm and parameters. The most accurate model was a single layer feed-forward neural network trained using monthly data. The CV(RMSE), NMBE and med(absRTE) for this model were evaluated to be -3.24%, 10.8% and 7% respectively. This represents a further 41.9% reduction in CV(RMSE) compared to that of the best performing model in the earlier analysis presented in Section 3.5.1. In addition to this, it is important to note that the spread of model error reduced as the measurement frequency increased. This is advantageous in developing consistently accurate models, as opposed to the sporadic performance at lower measurement frequencies.

The best performing models for the traditional and proposed approaches were brought forward to the analysis detailed in Section 3.5.3. For the proposed approach, this was the ANN constructed with monthly data and for the traditional M&V approach, this was the OLS regression model that contained just two predictor variables and used a monthly measurement frequency. The sensitivity of each model to the quantity of training and testing data available was investigated. It was found that the model constructed using the traditional approach was highly dependent on the length of the baseline period. Performance degraded across all testing data sets when the training period was reduced. In contrast to this, the ANN models were found to perform significantly better. Sufficient accuracy was achievable in all cases for training periods greater than 6 months. A CV(RMSE) of 10.8% was achieved with a 6-month training period and a 12-month testing period. This highlights the potential benefits of the proposed approach in overcoming the limitations of traditional M&V in industrial buildings. Interestingly, a 4-month training period and a 3-month testing period resulted in CV(RMSE) of 3.7%.

Chapter 4

A Machine Learning-Based Modelling Methodology

4.1 Introduction

4.1.1 Overview

The research detailed in Chapter 3 highlighted the benefits of utilising machine learning techniques for energy modelling in M&V applications. It has been noted previously that a review of different approaches must be carried out under a set of common conditions to allow for the best performing model to be identified (Zhao & Magoulès 2012). Thus, it was concluded that an exhaustive approach to modelling is needed for each application to ensure the optimal model is constructed. This is necessary as the characteristics of each algorithm are better suited to certain data sets than others. It is not possible to propose a single modelling technique for the broad spectrum of ECMs. While these techniques offer great opportunities for improvement in the energy efficiency field, they also present a problem to M&V practitioners. As they are not covered by the current guidance documentation, how can they be applied in a correct manner to ensure model accuracy, validity and reliability?

The foundations of all methodologies for the M&V of energy savings are based on the same six key principles: accuracy, completeness, conservatism, consistency, relevance and transparency. It has been noted that the most widely accepted methodologies tend to generalise M&V so as to ensure applicability across the spectrum of ECMs. These do not provide a rigid calculation procedure to follow. This chapter aims to bridge the gap between high-level M&V protocols and the practical application of machine learning modelling algorithms. The methodology was developed specifically for applications in the industrial buildings sector.

The development of a novel, machine learning supported methodology for M&V 2.0 enables accurate and reliable quantification of savings. A novel and computationally efficient feature selection algorithm and powerful machine learning regression algorithms are employed to maximise the effectiveness of available data. The baseline period energy consumption is modelled using ANN, SVM, k-NN and multiple OLS regression. Improved knowledge discovery and an expanded boundary of analysis allow more complex energy systems be analysed, thus increasing the applicability of M&V. A case study in a large biomedical manufacturing facility is used to demonstrate the methodology's ability to accurately quantify the savings under real-world conditions.

4.1.2 Background

The most widely recognised and well established methodologies for the M&V of energy savings were reviewed in Chapter 2 (Section 2.4). These include the IPMVP, ASHRAE Guideline 14 and ISO 50015. These methodologies are intertwined with one another and provide guidance on applying universal approaches to the wide spectrum of energy saving projects. Despite this, the lack of a rigid calculation process has been highlighted as a significant shortcoming of these protocols (Ginestet & Marchio 2010). This is less of an issue in residential and commercial applications as the nature of the energy systems in place are more simplistic. In contrast to this, industrial buildings contain complex energy systems with many variables impacting on energy consumption. ASHRAE Guideline 14 explicitly states that its procedures do not cover major industrial loads (American Society of Heating Refrigerating and Air-Conditioning Engineers 2014). The lack of a prescribed, analytical process that can be applied has implications on the accuracy and reliability of energy savings.

The major focal point of legislation has been the implementation of ECMs to minimise consumption in the industrial sector. For this to be a success, the M&V performed on each individual ECM must be of sufficient accuracy so that the savings estimated can be relied upon. The cumulative impact of these ECMs will be evidence of the success of the Directive. There is a significant danger that over estimation of savings on an individual project level could hinder attempts to limit climate change. This has created a need for a methodology that is capable of overcoming the barriers that impede accurate M&V in industrial facilities. These challenges include cost, resources and the time required to perform M&V. The potential of machine learning to aid the minimisation of uncertainty in M&V projects has already been assessed in Chapter 3. Building on the findings of this earlier analysis, this chapter outlines the development of a machine learning supported energy modelling methodology that is tailored to the characteristics of the industrial buildings sector. A clearly defined, prescriptive process focused on harnessing the power of energy data in an efficient manner is presented.

The integration of these advanced machine learning techniques, coupled with the use of larger data sets, presents new challenges to M&V practitioners. The critical advantages of traditional approaches to the problem lie in the simplicity of application. This is particularly useful in commercial buildings where the number of independent variables used to normalise consumption is small. In one case study of an office building, the OAT is sufficiently powerful to model the hot water consumption in the building (Zhang et al. 2015). In other such cases, monthly billing data is often used to create weather dependent OLS linear regression models that require very little computational resources. The machine learning-based methods proposed are in direct contrast to this. As discussed in Chapter 3 (Section 3.3), many of the more powerful knowledge discovery techniques can be computationally expensive. In addition, expert knowledge is often required to chose appropriate values for model hyper-parameters which directly contribute to prediction accuracy. Furthermore, while the availability of larger quantities of data enables the evolution of practices, it presents another barrier to the adoption of these machine learning techniques. How can one identify which data are useful for modelling the dependent variable of choice? This requires the use of feature selection techniques which are designed to identify relevant subsets of independent variables. Each of these challenges are addressed in the methodology presented with a focus on providing a robust, prescriptive approach that can be applied without requiring subject area expert knowledge. The methodology is designed to empower M&V practitioners in the use of machine learning techniques, while ensuring validity of the final baseline energy model constructed.

4.2 Research Questions

To date, AI has been proven to be advantageous in building energy load prediction. The primary objective of the research detailed in this chapter is the development of a replicable, robust and detailed methodology to enable the use of machine learning techniques for the purposes of M&V in industrial facilities.

Solutions to the following research questions were sought in carrying out this work:

1. Can a definitive methodology be developed to provide explicit guidance on the application of machine learning in M&V?
2. Is it possible for such a methodology to be robust enough to harness the power of available data across the spectrum of different M&V projects?
3. Can machine learning algorithms be employed on large data sets without increasing the resources required for M&V?
4. An extended boundary of analysis is proposed to increase the baseline energy model accuracy in circumstances with limited system specific metering infrastructure. Can M&V be completed with acceptable accuracy using this novel boundary of analysis?

In addition to these chapter specific research questions, RO2 and RO3 are the overarching motivators used to guide the research presented in this chapter.

4.3 Feature Selection

The challenge of identifying subsets of data which are relevant and beneficial to the construction of a baseline energy model was highlighted in Section 4.1.2. Feature selection techniques offer an effective solution to this problem, while also offering the benefit of decreasing model training time by identifying a subset of useful independent variables and discarding the remaining data which add noise to the training process. Feature selection is otherwise known as variable selection or attribute selection. It is the automatic selection of attributes that are most relevant to the predictive modelling problem in question. The primary objectives of the process are to improve the prediction performance of the predictors, provide faster and more cost-effective predictors, and present a better understanding of the underlying process that generated the data (Guyon & Elisseeff 2003). There are three general classes of feature selection algorithms: filter methods, wrapper methods and embedded methods.

The overall problem can be synthesised in the following manner. F is the given set of original features or independent variables (i.e. the raw data set) with cardinality n, where n symbolises the number of features in set F. $\overline{F}$ is the selected subset of features with cardinality $\overline{n}$, where $\overline{n}$ represents the number of features in the subset and $\overline{F} \subseteq F$. Finally, $J(\overline{F})$ is the selection criterion applied and it is assumed that a higher value of $J(\overline{F})$ indicates a better feature subset. With the objective of maximising $J(\overline{F})$, the problem is defined as follows (Wang et al. 2009):

$$J(\overline{F}) = \max_{Z \subseteq F, |Z|=n} J(Z) \tag{4.1}$$

A highly cited publication by Guyon & Elisseeff (2003) establishes the following ten step process that should be adhered to in solving any feature selection problem:

1. Is domain knowledge available? If yes, it is advisable to construct an *ad hoc* feature set using this knowledge. It should be noted that this is often difficult when dealing with complex energy systems in industrial facilities.

2. Do the variables correspond in size? If no, they need to be normalised. This eliminates the units of measurement and enables easier comparison of data from different sources.

3. Is interdependency in the data set suspected? The dependence of independent variables on both the dependent variables and each other can introduce error into a model. It is recommended that the feature set is expanded by constructing conjunctive features or products of features.

4. Do the input variables require pruning? If no, disjunctive features or weighted sums of features should be constructed.

5. Do features need to be analysed independently? This is completed in cases where the impact of individual features needs to be quantified. If yes, a variable ranking method should be used. It is recommended that this step is completed even if not explicitly required.

6. Is a predictor or independent variable required? If no, stop.

7. Is the quality of the data poor? Dirty data exists when noise or meaningless data are included in the data set. If yes, outliers should be detected using the top ranking variables. Unclean data should be discarded to maximise model performance.

8. Is the most suitable feature selection algorithm known? If no, utilise a linear predictor. Use a forward selection method or the ℓ_0-norm embedded method. This should then be compared with a sequence of ranked predictors using increasing subsets of features. Can performance be matched or

improved with a smaller subset? If yes, try a non-linear predictor with that subset.

9. Are new ideas, time, computational resources and sufficient data available? If yes, compare several different feature selection methods.

10. Is a stable solution desired? If yes, utilise bootstrapping methods.

4.3.1 Filter Methods

Filter methods identify the intrinsic properties of the data set using univariate statistics. These approaches are fast, scalable and independent of the modelling algorithm. However, they ignore feature dependencies as each variable is assessed independently. The interaction with the final modelling algorithm is also ignored (Saeys et al. 2007). While some features are statistically significant when reviewed individually, they may not be well suited to the characteristics of the modelling algorithm. Examples of the metrics used to select features include the information gain, chi-square test, fisher score, correlation coefficient and variance threshold.

4.3.2 Wrapper Methods

Wrapper methods embed the modelling algorithm in the feature selection process (Saeys et al. 2007). They are the most popular approach to feature selection (Wang et al. 2009). A set of feature subsets is defined to begin with. Each subset is then trained on the modelling algorithm and evaluation enables the identification of the optimal subset (Kumari & Swarnkar 2011). The benefits of such approaches include simplicity, interaction with the learning algorithm, assessment of interdependencies and accuracy. However, it is imperative to note that these approaches are more computationally expensive than filter-based approaches. Also, wrapper methods themselves can be sub-classified as either deterministic or randomised algorithms. Recursive feature elimination, sequential feature selection algorithms and genetic algorithms can all be classified as wrapper methods.

4.3.3 Embedded Methods

Embedded methods learn which features best contribute to the accuracy of the model while the model is being created. The most common type of embedded feature selection methods are regularization methods. Thus, embedded methods are specific to a given learning algorithm (Saeys et al. 2007). Benefits include the direct interaction with the modelling algorithm, improved computational complexity compared to wrapper methods and assessment of model feature dependencies. A key drawback is the selection process being based on a specific learning algorithm. Examples of regularization algorithms are the LASSO, Elastic Net and Ridge Regression.

4.4 Methodology

AMI is common place in modern industrial buildings. These systems collect large quantities of granular energy data, which can be used to discover knowledge on system behaviour. This data is widely available, however, it is rarely utilised to its full potential. Common issues that hinder the use of this data include the lack of a central data repository, inefficient preprocessing techniques and insufficient subject matter knowledge. This data can significantly improve the accuracy with which M&V can be performed, although the need for skilled professionals to perform tasks such as data cleaning and baseline energy modelling impedes the process.

The methodology presented in this chapter is capable of overcoming the issues impeding the effective use of available data. A novel alternative to the traditional M&V protocol is offered. In contrast to the IPMVP and ASHRAE Guideline 14, a prescriptive data handling and modelling procedure is detailed to ensure that maximum accuracy is achieved. In addition to this, the prescription of this methodology contributes to simplified decision making and reduces the need for subject matter expertise.

It is important to note that the use of additional modelling techniques, not included in this methodology, is at the discretion of the M&V practitioner. Additionally, the data handling framework is not completely prescriptive so as to maintain applicability and remain largely technology agnostic. Detailed data pipelines have been developed for data-driven analytics applications in large-scale industrial facilities and these can be integrated into the proposed methodology by the user (O'Donovan et al. 2015). Chapter 2 (Section 2.5.4) details this further. Figure 4.1 illustrates the process flow diagram of the methodology.

4.4.1 Step 1 - Definition of Project Parameters

It is critical that the scope and boundaries of the project are defined prior to any commencement of work. This ensures that the M&V of the resultant energy savings can be completed in an accurate, complete, conservative, consistent, relevant

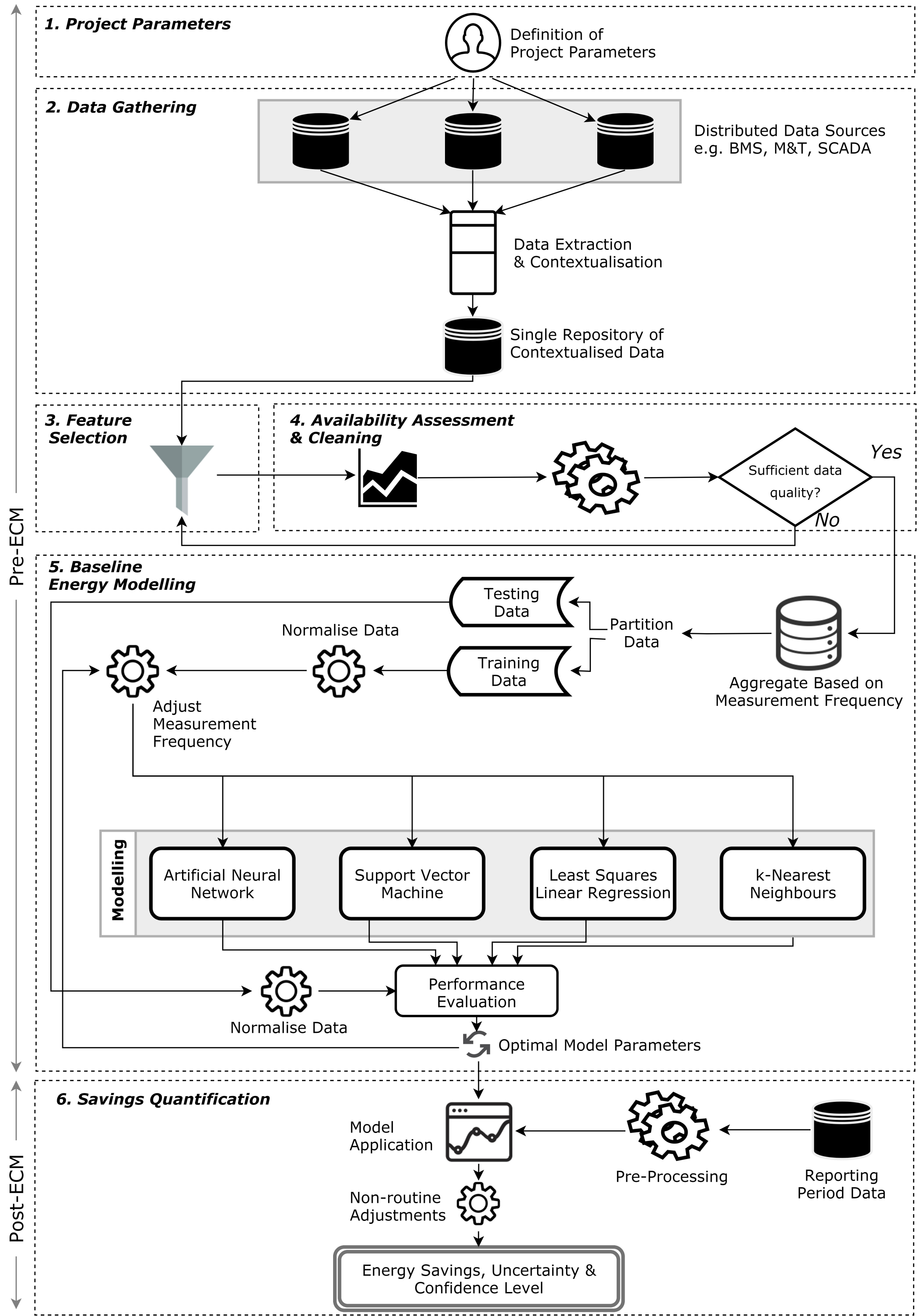


Figure 4.1: Process flow diagram of the proposed baseline energy modelling methodology

and transparent manner. To do this, the following items need to be documented;

- ECMs to be carried out
- Project boundary
- Relevance to total energy consumption on entire site
- Project time-line: Expected baseline, implementation & reporting periods
- Relevant personnel
- Data sources
- Static factors

4.4.2 Step 2 - Data Gathering

Extraction

The characteristics of each relevant data source identified must be detailed. This should include the type of data, measurement frequency, storage methods and access protocol. The objective of this stage in the process is to outline a means of accessing data from each distributed data source to enable data extraction. Information on how to manage and access this data beyond the baseline period must also be included. As per ISO 50015, this includes, but is not limited to, storage, backup, maintenance and security of the data. The information collated at this stage should be replicable so that the same process can be followed during the reporting period.

Contextualisation

Contextualisation of the relevant data is critical in gaining meaningful insights into the systems being analysed. Poor semantic modelling is common for energy data across the industrial sector. This has led to the need for a standardised methodology for describing data. One such solution to this problem is Project Haystack (*Project Haystack* 2018). The goal of the Haystack naming convention is to make it easier and more cost effective to analyse, visualise and derive value from operational data. The object oriented class hierarchy, illustrated in Figure 4.2, is based on three entities: the site, pieces of equipment and points. The naming convention uses a tag model to describe data within the context of the facility. A full reference guide for applying the naming convention is available online (*Project Haystack* 2018).

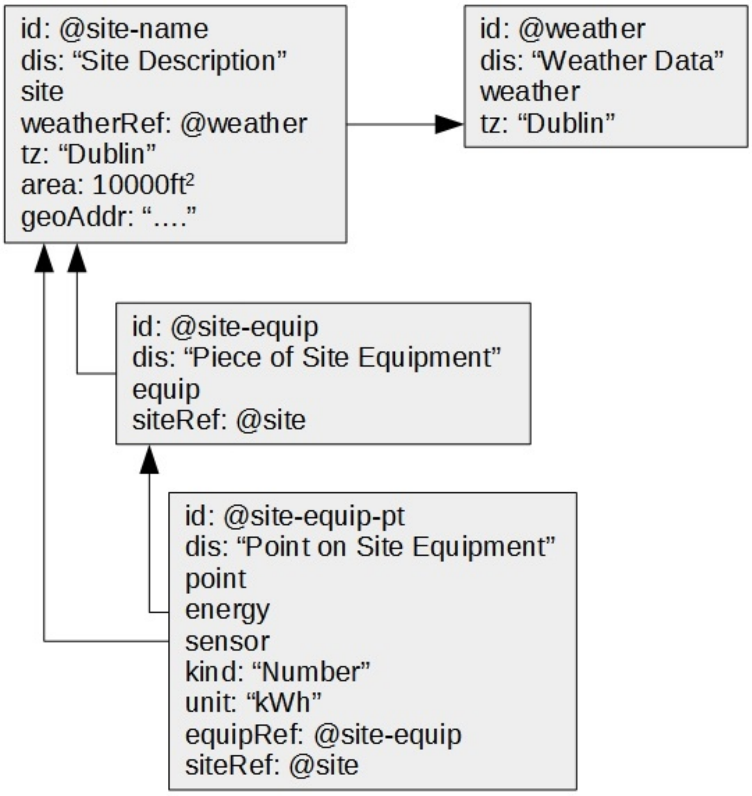


Figure 4.2: Basic three level hierarchy of the Haystack naming convention

The Haystack naming convention is endorsed in this methodology as it has many uses far beyond the M&V application. If deployed across systems on a site, communication and data accessibility is naturally improved, and hence analytics, is made more straightforward. In the context of this methodology, other naming conventions can be applied if desired. This step is important to ensure a contextualised, more useful data set is produced.

4.4.3 Step 3 - Feature Selection

The contextualised data set can often be very large as hundreds of variables are stored. It is important that only those variables that offer importance in model construction are brought forward for analysis. Therefore, feature selection is used to select a subset of relevant variables for use in model construction.

To clarify, the term "variable" is used to refer to the raw input data and the term "feature" is used for those variables output from the feature selection process. These will be the features used at a later stage to construct the baseline energy consumption models. Feature selection has many benefits in this application, including reducing the measurement and storage requirements in the reporting period, minimising model training time and avoiding high dimensionality for improved prediction performance (Guyon & Elisseeff 2003).

A combination of simple filter and wrapper approaches are employed in this methodology to minimise processing speed, while selecting the optimal feature subset. A Spearman rank correlation filter method is included in a wrapper that seeks to maximise the adjusted coefficient of determination ($R^2_{adjusted}$) of the data set. This is found by constructing a multiple OLS model. This approach measures the strength and direction of monotonic association between two variables using Algorithm 1. The Spearman correlation coefficient is defined as the Pearson correlation coefficient between two ranked variables. This is then used to iteratively select variables to add to the feature subset. The addition of each feature is evaluated with respect to the $R^2_{adjusted}$ value produced using the previous feature subset. Algorithm 2 details this iterative process which determines the optimal subset of features. This approach was developed as the benefits of both the wrapper and filter-based methods can be realised, while minimising computational speed through the use of a simple OLS regression model. The use of more advanced algorithms such as ANN and SVM would act as a barrier to adoption due to their significantly larger computational requirements.

To avoid multicollinearity, it is critical that the features identified are not just independent of the dependent variable, but also independent of each other. Multicollinearity can cause coefficient estimates in multiple regression models to change erratically in response to small changes in the model or data. The variance inflation factor (VIF) is used to test for multicollinearity between features and hence, avoid selecting redundant features. Similar to the coefficient of determination, there is no single value for the VIF that indicates multicollinearity. A commonly used rule of thumb is a VIF value greater than 10 indicates multicollinearity. For weaker models, values above 2.5 may cause concern. In keeping with the principle of conservatism applied in the case of the $R^2_{adjusted}$, any feature found to have a VIF greater than 5 should be removed from the feature set.

4.4.4 Step 4 - Availability Assessment and Cleaning

A data availability assessment consists of an initial, high-level statistical analysis of the proposed model features. The results of this allow the practitioner to make an informed decision, based on data quality and integrity, as to which features are suitable for analysis. The simple summary statistical measures detailed in

Algorithm 1: Calculation of Spearman rank correlation

Input: Input data set expressed as an m x n matrix.
$x[\ ,n]$ = dependent variable
for $j = 1, \ldots, n-1$ **do**
  r_j = rank x[,j]
  r_n = rank x[,n]
  if *All m ranks are distinct integers* **then**
    for $i = 1, \ldots, m$ **do**
      Compute
      $d_i = r_j[i] - r_n[m]$
    end
    Compute
    $\rho = 1 - \frac{6\Sigma d_i^2}{m(m^2-1)}$
  else
    for $i = 1, \ldots, m$ **do**
      Compute $\rho = \frac{cov(r_j, r_n)}{\sigma_{r_j}\sigma_{r_n}}$, where
      $cov(r_j, r_n)$ is the covariance of the rank variables
      $\sigma_{r_j} \& \sigma_{r_n}$ are the standard deviations of the rank variables
    end
  end
end
Output: Spearman correlation coefficients, ρ, for each input variable w.r.t. the dependent variable.

Table 4.1 are to be used to enable evidence based decision making.

Features with large numbers of outliers, periods of missing data or unreliable measurements should be omitted from the feature subset. As a rule of thumb, features with more than 5% of poor quality data should be omitted entirely from the subset. Any features that fall short of this 5% omission threshold can generally be cleaned using the process detailed in Section 4.4.4. In addition to the summary statistics, visualisation techniques can also be used to gain an understanding of the data at hand. Box plots, time series plots and histograms are useful in graphically representing the data. This process ensures that data quality and integrity is maintained. Section 4.5 presents an implementation of this assessment.

Data cleaning is the process of detecting and removing inaccurate entries in a data set. Maintaining quality in the baseline period data is critical to ensuring the system under analysis is accurately modelled. Under the IPMVP, baseline data should not be replaced by modelled data, except when using Option D (Efficiency Valuation Organization 2012). Therefore, the scope of data cleaning

Algorithm 2: Spearman rank-based feature selection to optimise adjusted coefficient of determination

Input: m x n matrix containing all input data.
$x[\ ,n]$ = dependent variable
Apply algorithm 1 to calculate variable *ranks*
Order columns in input matrix by decreasing ρ
$i = 1$
ρ_i = Spearman correlation coefficient between variable i and $x[,n]$
$subset_i = x[\ ,cols(1,\ldots,i,n)]$, i.e. variable with highest ρ and dependent variable
Train OLS regression model for $subset_i$ & find $r^2_{adj_i}$
while *i != no. of variables* **do**
    $subset_{i+1} = x[\ ,cols(1,\ldots,i+1,n)]$
    Train OLS regression model for $subset_{i+1}$ & find $r^2_{adj_{i+1}}$
    if $r^2_{adj_{i+1}} - r^2_{adj_i} > 0.01$ **then**
        $subset_i = subset_{i+1}$
        $r^2_{adj_i} = r^2_{adj_{i+1}}$
        $i = i + 1$
    end
    else
        Remove variable $i + 1$ from the data set
    end
    return $subset_i$
end
Output: data set with features selected.

Table 4.1: Statistical measures to be employed in the data availability assessment

Measure	Description
Mean	The average value in a set of numbers
Median	The value lying at the midpoint of a frequency distribution of values.
No. of Unique Values	The number of unique values in the set of measures for a variable.
No. of Missing Values	The quantity of values missing the data set. This is usually assessed with respect to a measurement frequency being used in the analysis.
Quartiles	The three points that divide the data set into four equal groups, each group comprising a quarter of the data, where the data is ordered sequentially.
Minimum	The lowest value in the set.
Maximum	The highest value in the set.

in this application is limited to simply identifying unclean data and subsequently removing it. No backfilling of data is to take place. The only exception to this is if data is missing for a consistent period of time in the baseline period, then comparable data for the same time period in a different year can be employed.

The results of the data availability assessment carried out in Section 4.4.4 are to be used to guide the data cleaning process. Using the summary statistics, box plots, time series plots and histograms output from the assessment, variables with irregularities can be identified. Some of these irregularities cannot be rectified with data cleaning; these features must be omitted from the analysis. If features must be omitted, the feature selection algorithm should be reapplied with these features removed from the data set. This can allow for new features to be included in the data set.

4.4.5 Step 5 - Baseline Energy Modelling

The development of the optimal model of the baseline energy consumption is critical to ensuring that the uncertainty associated with the final energy savings is minimised. As stated already, this model is referred to as the baseline energy model. An exhaustive process is used to ensure the model developed is tailored to the characteristics of each specific project.

Aggregate Based on Measurement Frequency

Firstly, the data is aggregated based on the measurement frequency. This is necessary as each subsequent step is frequency specific. The objective of this step is to generate multiple data sets to enable an array of models be developed. The number of data sets that can be created is dependent on the frequency with which the data is measured. The wide availability of AMI in modern industrial buildings generally results in data being recorded in 15-minute intervals. For this case, the data is then aggregated using the mean values for hourly and daily measurement frequencies. This results in 3 data sets being available for model development purposes. It is not advised that less granular data than that with a weekly measurement frequency be used as these can result in insufficient quantities of testing data leading to unreliable results.

Partitioning of Data

The data gathered at this stage of the process is for the baseline period only as the ECM has not yet been implemented. The data sets output from the previous aggregation stage are partitioned into training and testing data sets. This enables the models to be constructed using the training data and tested on an unseen set. A shuffled split is performed with 80% of data used for training and 20% used for testing. This is in contrast to the guidance given by the IPMVP which uses 100% of the baseline data to construct the model and subsequently calculates the performance metrics by applying the model to the same data set (Efficiency Valuation Organization 2014). In cases with multiple degrees of model freedom, this approach may be prone to over-fitting the model to the training data, which can result in random error or noise being incorporated, resulting in unreliable performance evaluation. The partitioning of the baseline period data, and testing on a data set not used in model construction may prove a more accurate approach that results in a reliable and independent evaluation of performance.

The training data is brought forward to the next stage in the methodology, while the testing data set is not used again until performance evaluation is required in Section 4.4.5.

Feature Scaling

Feature scaling is used to standardise the range of features in the data set with a view to improving model performance. It also improves the processing time of certain algorithms including ANNs. This process is also known as standardisation or Z-score normalisation and results in each feature having the properties of a standard normal distribution (i.e. standard deviation of 1 and mean of 0). Each of the training data sets input into this stage of the process are standardised using Equation 4.2 and the scaling parameters of each feature in these data sets are stored for application at a later stage.

$$Z = \frac{x_i - \mu}{\sigma} \tag{4.2}$$

where, x_i is the value being standardised, μ is the mean of the distribution and σ is the standard deviation of the distribution.

Model Training

Baseline energy models are trained using the data sets for each measurement frequency. This is an exhaustive process which seeks to identify the most appropriate model hyper-parameters for each algorithm and measurement frequency. Chapter 2 reviews the success of various machine learning algorithms in the field of M&V to date. The ability of machine learning algorithms to model energy systems in industrial buildings outside the context of M&V is also presented. The findings of this review and the results of the analysis detailed in Chapter 3 led to the following algorithms being deemed the most appropriate for application:

1. Ordinary least squares regression (OLS)
2. k-Nearest neighbours regression (k-NN)
3. Artificial neural network (ANN)
4. Support vector machine regression (SVM)

Each of the algorithms requires the values of certain parameters be set prior to the commencement of the learning process. These are known as hyper-parameters. A grid-search approach using 10-fold cross validation is employed to find the optimal values of each hyper-parameter. The recommended grid-search values for each hyper-parameter are detailed in Table 4.2. As this is not an exhaustive list of grid-search values, practitioners may choose to alter the specifications of each. The values provided are recommended based on previous research on the success of those employed in Chapter 3. The training of models is an optimisation solution which uses an iterative approach to arrive at the final values of the hyper-parameters.

A model constructed using each algorithm and measurement frequency is output from this stage. For a case assessing 3 measurement frequencies, there are 12 models developed.

Performance Evaluation

To identify the most suitable model, the testing data is used to evaluate the performance of each baseline energy model. The testing data sets defined in Section 4.4.5 are standardised using the scaling factors employed on the training data sets. These are specific to each modelling frequency. Each model developed is then applied to the appropriate standardised data set to produce a prediction of energy consumption.

The CV(RMSE) and NMBE are employed to quantify the prediction performance of each model. CV(RMSE) is a commonly employed performance metric and is used in both IPMVP and ASHRAE Guideline 14. It is a measure of the variability between actual and predicted values. The CV(RMSE) is employed as it gives context to the size of the error relative to the quantity being modelled. It is also

Table 4.2: Description of algorithms and associated hyper-parameters

Algorithm	Description	Hyper-parameters	Grid Search
Bi-variable Linear Regression	An ordinary least squares approach assumed to be representative of typical M&V practice.	Intercept	True/False
Multi-variable Linear Regression	A more detailed ordinary least squares model constructed using 9 additional features from the available data set.	Intercept	True/False
k-Nearest Neighbours	Non-parametric model where the input consists of the k closest training examples in the feature space. The output is the average of the values of its k-nearest neighbours.	Maximum no. of neighbours Distance Kernel	k_{max} = 1:10 d = 1:5 $kernel$ = triangular
Artificial Neural Networks	Non-linear statistical model. It is a two-stage regression model typically represented by a network diagram. A single hidden layer feed-forward neural network was developed in each instance.	No. of hidden units Maximum no. of iterations Threshold Weight decay	$size$ = 1:10 it_{max} = 1,000 t = 0.01 d = (0.001,0.01, 0.1,0.5)
Support Vector Machines	Non-parametric technique reliant on kernel functions. Examples are represented as points in space with a clear gap separating mapping categories.	Kernel Cost	$kernel$ = linear c = (0.25,0.5,1)

important to note that the RMSE is known as the SE in the IPMVP.

NMBE is an indication of overall bias in a regression model. It quantifies the tendency of a model to over or under-estimate across a series of values. In contrast to the CV(RMSE), the NMBE is independent of time and hence, it can result in overall positive bias cancelling out negative bias. The use of both metrics in conjunction with each other allows for a true insight into model performance.

Equations 4.3 and 4.4 are used calculate each metric, where y_i is the actual value, $\hat{y}_i$ is the predicted value, $\bar{y}$ is the average of the actual value, and n is the total number of predictions in the period of analysis.

$$CV(RMSE) = \frac{1}{\bar{y}} * \sqrt{\frac{\sum_i^n (y_i - \hat{y}_i)^2}{n-k-1}} * 100 \tag{4.3}$$

$$NMBE = \frac{1}{n-1} * \frac{\sum_i^n (y_i - \hat{y}_i)}{\bar{y}} * 100 \tag{4.4}$$

The best performing model is identified as the model that results in the lowest CV(RMSE), as this is the metric used to calculate the uncertainty introduced. The NMBE is used to support the model evaluation, however, it is not used in model selection due to the possibility of positive bias cancelling out negative bias and the objective of minimising modelling uncertainty.

4.4.6 Step 6 - Savings Quantification

The previous steps are all completed prior to the implementation of the ECM. This ensures that any shortcomings in the approach are identified at an appropriate time such that remedial actions can be taken. For example, if the feature selection algorithm showed that the variables were not strongly correlated to the dependent variable, then additional metering would be required to gather the necessary data in the baseline period. With these steps completed, the implementation period is used to implement the ECM and perform any commissioning works that are necessary. The final energy savings can then be quantified so long as the necessary reporting period data is available.

Data Gathering

Data for the final subset of model features must be gathered to enable calculation of the adjusted baseline. The information for data sources and associated data points logged in Section 4.4.2 is used to guide this process. The data for both the model features and static factors must be gathered for the entirety of the reporting period.

Preprocessing

The data gathered must be preprocessed to ensure integrity and accuracy is maintained. This preprocessing involves transforming the data set into a format suited to the baseline energy model. The contextualised feature names, measurement frequency and cleaning are all covered in this stage. As per ASHRAE Guideline

14, independent variables must not be more than 110% of the maximum and no less than 90% of the minimum values of the corresponding model training data. If the data does not conform to this requirement, an advisory note must be attached to the savings to state that the data is beyond the range of applicability of the model (American Society of Heating Refrigerating and Air-Conditioning Engineers 2014).

As the model is trained on a standardised data set, all input data must be standardised to the same scaling parameters. Therefore, the reporting period data must be scaled to the same mean and standard deviation as the training data set. This standardisation is required to maximise the performance of the modelling algorithms.

Model Application

The optimal model identified in Section 4.4.5 is applied to the prepared reporting period data to calculate the adjusted baseline. The z-score normalisation performed on the reporting period data must now be reversed to give a context to the adjusted baseline. This can then be directly compared to the measured data for the same period of analysis.

Non-routine Adjustments

A non-routine adjustment is required when changes to static factors that affect the consumption of an energy system occur. These are typically changes to a facility's operations, size or equipment. The baseline energy model is used to account for all routine adjustments. In contrast, manual modifications to an M&V methodology are required to make non-routine adjustments. They are implemented on a project-by-project basis in circumstances with changes in static factors. Each non-routine adjustment is a custom engineering calculation for the given problem. They must be agreed upon by all project stakeholders.

Uncertainty

The energy savings estimated must have an associated level of uncertainty and confidence. It is important to note that differing approaches for the calculation of uncertainty are proposed by the IPMVP and ASHRAE Guideline 14. In the IPMVP, acceptable uncertainty requires the savings to be larger than twice the standard error of the baseline value (Efficiency Valuation Organization 2014). ASHRAE Guideline 14 states that uncertainty must be less than 50% of the annual reported savings, at a confidence level of 68% (American Society of Heating Refrigerating and Air-Conditioning Engineers 2014). The approach deemed suitable by the IPMVP is employed in this methodology. Thus, Equation 4.5 is applied, where t is the t-statistic for a given level of confidence and degrees of freedom and SE is the standard error of the estimate.

$$U = t * SE \tag{4.5}$$

In the IPMVP, standard error is calculated using all of the baseline data. As discussed previously in Section 4.4.5, this technique of using all baseline data to train and test the model is prone to over-fitting the model to that specific data set. The performance metrics are in turn calculated based on the models ability to fit the baseline data and subsequently, uncertainty is calculated. This can result in low levels of model error on the baseline data, but unreliable measures of uncertainty in the reporting period. The introduction of a random data split overcomes these issues by applying the model to an unseen testing data set.

The ASHRAE approach to calculating uncertainty is not employed as it is too susceptible to the size of the baseline data set. It is found in a similar fashion to the IPMVP approach using the CV(RMSE) in Equation 4.6, which assumes that there is zero error introduced by the metering equipment; thus only the uncertainty introduced by model is calculated. The quantity of data available, required reporting period length and CV(RMSE) are the only variables influencing the quantity of uncertainty introduced in this phase. It is clear in this equation that as the quantity of training data available in the baseline period increases, the uncertainty reduces. The same is true for the length of the reporting period. To meet the uncertainty requirements already discussed, the required model accuracy reduces as the measurement frequency increases and the importance of an accurate baseline model is diminished. This is seen as a flawed approach that does not exude confidence in results.

$$U = t * \frac{1.26 * CV(RMSE)}{F} * \sqrt{\frac{n+2}{n*m}} \tag{4.6}$$

The model performance required to ensure acceptable uncertainty changes relative to the quantity of savings resulting from an ECM. This relationship, based on the IPMVP approach, is illustrated in Figure 4.3 and can be used as a reference chart for establishing the required performance levels.

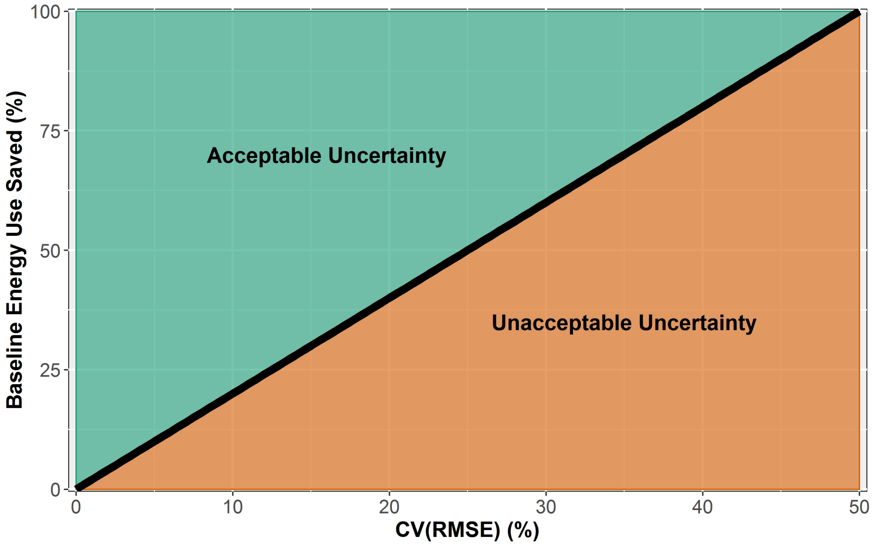


Figure 4.3: Model prediction performance requirements under varying fractional savings

4.5 Case study: Application and Results

The proposed methodology was applied to quantify the savings resulting from an ECM carried out on the chilled water system in a large biomedical manufacturing facility in Limerick, Ireland. The facility operates a continuous production process on a 24/7 basis. The generation and distribution of chilled water consumes approximately 7-8% of the total site energy consumption annually. The chilled water generated by a series of electrically powered chillers is used to satisfy the space cooling loads across the facility. This is delivered using an array of AHUs. The ECM consisted of the reduction of chilled water consumption at an end-use level across the facility. This was achieved by identifying and optimising any AHU that were operating outside of design specifications. Figure 4.4 illustrates the energy consumption of the chilled water system prior to the implementation of the ECM, i.e. the baseline energy consumption.

The proposed approach was deemed suitable due to a number of project specific constraints. The whole-building approach outlined by the IPMVP was not suitable as the savings were estimated to be considerably less than 10% of the total sites consumption. In addition to this, there was insufficient metering infrastructure to allow the successful application of the standard retrofit isolation Options A or B. In residential and commercial buildings, it is often appropriate to utilise outside air temperature as the independent variable for modelling chilled water consumption, however, in this case the energy system has added complexity due to the production process in operation. The relationship between chilled water

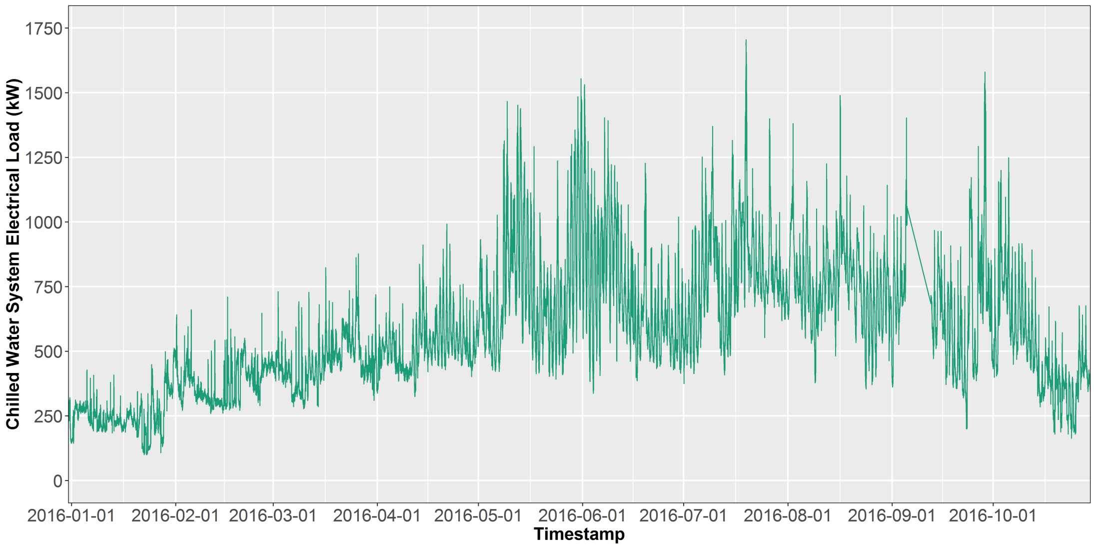


Figure 4.4: Electrical load of chilled water system in baseline period (pre-ECM)

system electricity consumption and cooling degrees days was found to be weak with a R^2 of 0.36. In addition to this, the lack of availability of granular production data restricted analysis into the relationship between production output and chilled water consumption. To implement a conventional approach, additional metering would need to be installed, which in turn would delay the project implementation as baseline data would have to be gathered.

Hence, the proposed methodology offers a novel alternative to the traditional approaches of M&V. The methodology can be used to establish the relationships between total electrical consumption of the chilled water system and a variety of other metered quantities on site. These metered quantities act as proxies of the production activity in the facility. Any significant relationships with independent variables can then be used to model the energy consumption in the baseline period, with a view to predicting the adjusted baseline following the implementation of the ECM.

4.5.1 Step 1 - Definition of Project Parameters

Prior to the commencement of any ECM implementation works, the following project scope and parameters were defined:

- **ECM**: Optimisation of AHUs to minimise the consumption of chilled water. This results in meeting the space cooling load with an increased efficiency.
- **Boundary**: The ECM will result in savings being achieved in the chilled water system electricity consumption. All other secondary benefits are out-

side the scope of this analysis.

- **Relevance to total facility consumption**: Chilled water system accounts for approximately 7-8% of site electricity consumption.
- **Baseline period**: 1st January 2016 to 29th October 2016.
- **Implementation period**: 30th October 2016 to 15th February 2017.
- **Reporting period**: 16th February 2017 to 25th September 2017.
- **Relevant personnel**: Facilities engineering team, M&V practitioner.
- **Data sources**: building management system (BMS), energy management system (EMS).
- **Static factors**: Number of production lines, size of facility, production process, shift scheduling, building fabric, space heating and cooling setpoints, air change rates.

4.5.2 Step 2 - Data Gathering

The existing metering infrastructure was utilised to develop a model for the chilled water system electricity consumption. Both the BMS and EMS store valuable data gathered by electrical, mechanical and climatic meters located across the facility. The characteristics of each data source are documented in Table 4.3. This will be referred to in the reporting period to replicate the data gathering process.

Table 4.3: Characteristics of data sources

Characteristic	BMS	EMS
Type of data	Mechanical, electrical & climatic	Electrical
Measurement frequency	Varies	15-min
Storage	On-site server	Remote server
Access	Local network access	Cloud access

As discussed in Section 4.4.2, poor semantic modelling of energy data in industrial facilities is common. The Haystack naming convention was applied to the data set to ensure each data point has a context with regard to the site. This increases the ease with which the data-driven model is then applied to data gathered in the reporting period. Figure 4.5 gives an example of the application of the Haystack naming convention to a sample data point. The two data sources were then collated together into a single data set of contextualised data points.

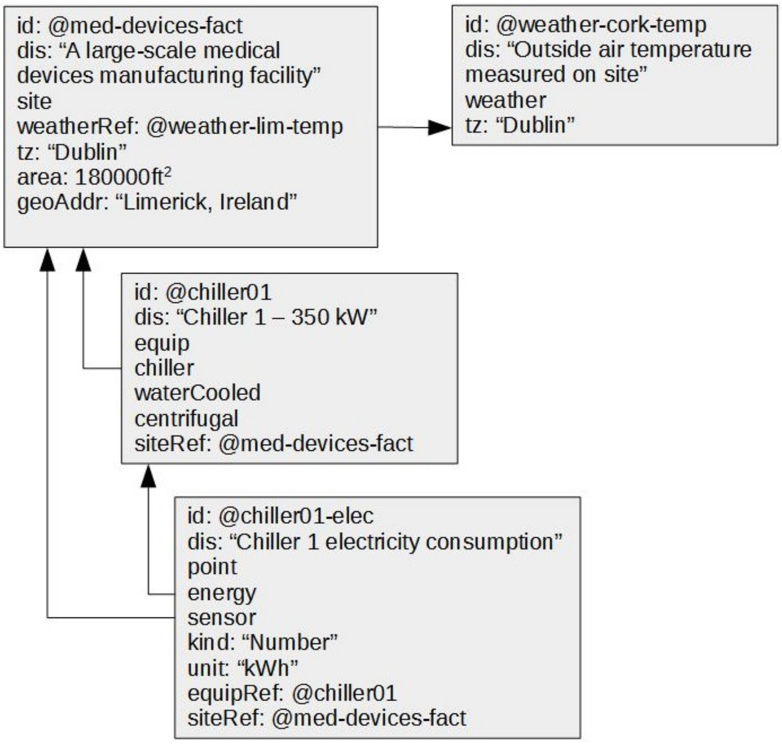


Figure 4.5: Example of the application of the haystack naming convention to the data set.

4.5.3 Step 3 - Feature Selection

The data set available for analysis contained 505 variables, each corresponding to a unique physical meter on-site. When the dependent variable, the chilled water system electricity consumption, is removed from this set, there are 504 variables that can be input to the model development process. The use of all of these variables to construct a model of the dependent variable would not be sensible given the computing resources typically available to practitioners. All variables used to model the consumption must add significant value to the model to be considered statistically significant.

Application of the feature selection algorithms outlined in Section 4.4.3 were applied to the data set. This resulted in the identification of 15 variables that added value to the multiple regression model explaining the dependent variable. Collinearity tests were conducted to ensure all features are independent variables and multicollinearity (interdependency) did not exist.

4.5.4 Step 4 - Availability Assessment and Cleaning

Box and whisker plots were used to visualise the statistical measures detailed in Section 4.4.4. These plots are included in Figure 4.6. Any points plotted outside the whiskers are considered outliers. Outliers are classified as being greater or less than 1.5 times the inter quartile range.

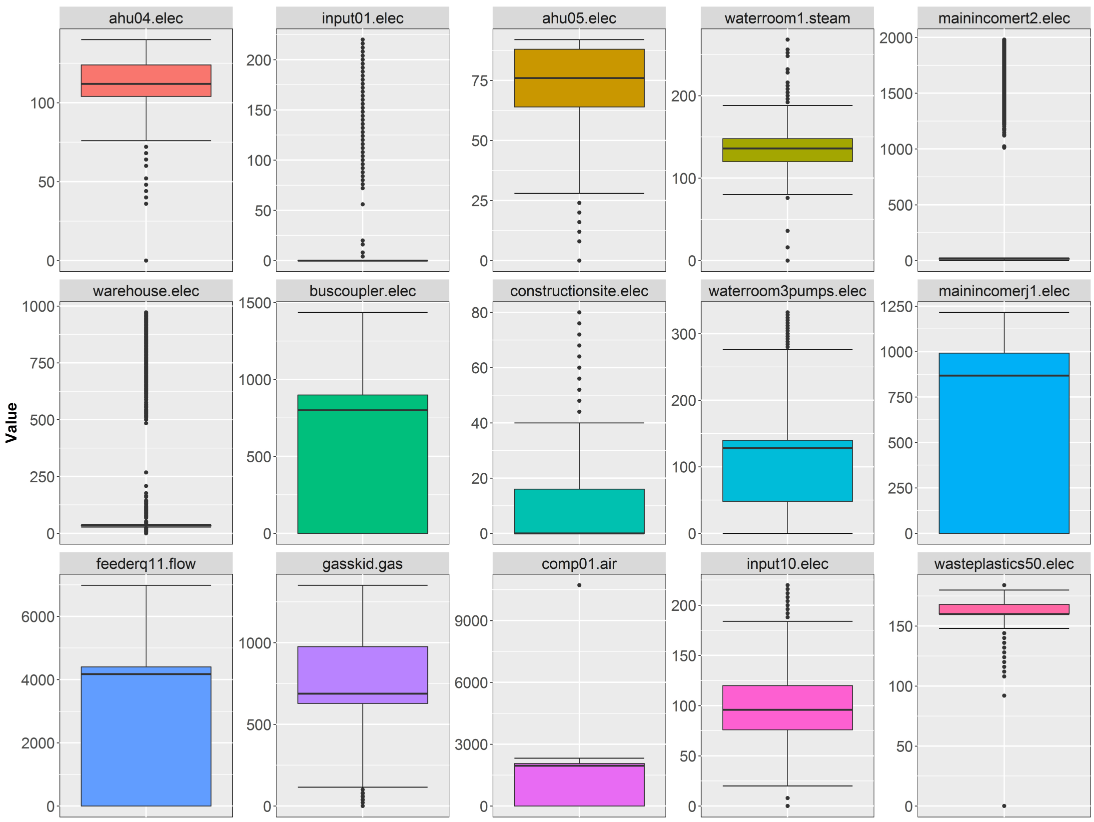


Figure 4.6: Box and whisker plots generated to evaluate each proposed model feature.

The results of the availability assessment are used to identify which features require cleaning. To comply with the IPMVP practices, this cleaning simply consists of omitting features identified as unclean from the analysis. No back-filling of unclean data is permissible. Any feature that had more than 5% of data identified as outliers were omitted from the feature set.

As a result of the data cleaning, 10 features were identified as being suitable for baseline energy model training. These features, or independent variables, selected to model the chilled water system electricity consumption are detailed in Table 4.4. For confidentiality reasons, the fully contextualised feature names have not been included.

The $R^2_{adjusted}$ value for these 10 features was 0.663. Although no strict limit for

Table 4.4: Feature set output from application of feature selection algorithm

Feature	Description
ahu04-elec	Electricity consumption of AHU no. 4.
ahu05-elec	Electricity consumption of AHU no. 5.
constructionsite-elec	Electricity consumption of construction site.
waterroom3pumps-elec	Electricity consumption of pumps in water room no. 3.
mainincomer-elec	Electricity consumption recorded on site incomer from grid.
feederq11-elec	Production related electricity consumption.
gasskid-gas	Gas consumption of gas skid.
comp01-air	Compressed air produced by compressor no. 1
input10-elec	Production related electricity consumption.
wasteplastics50-elec	Electricity consumption related to processing of production waste.

acceptable $R^2_{adjusted}$ exists, this value is on the lower end of acceptable values. This value does however increase as measurement frequency decreases indicating that the relationship is stronger as data granularity increases. This could possibly be due to delays between variable responses that do not cause an issue with lower measurement frequency.

4.5.5 Step 5 - Baseline Energy Modelling

The initial data set gathered has been minimised to include the chosen independent variables, or features, and the dependent variable; this is known as the feature set. The feature set was gathered using a 15-minute measurement frequency. The data was aggregated to create four independent feature sets with 15-minutes, hourly, daily and weekly measurement frequencies. This enables an exhaustive approach to modelling, which results in the identification of the optimal data granularity.

Each feature set was then split into training and testing data using an 80:20 ratio. This allows models to be evaluated on unseen data, hence, improving reliability of results. The training data was then standardised to ensure the best possible model fit is achieved. An OLS, k-NN, ANN and SVM regression model was trained for each measurement frequency. A grid search approach was used

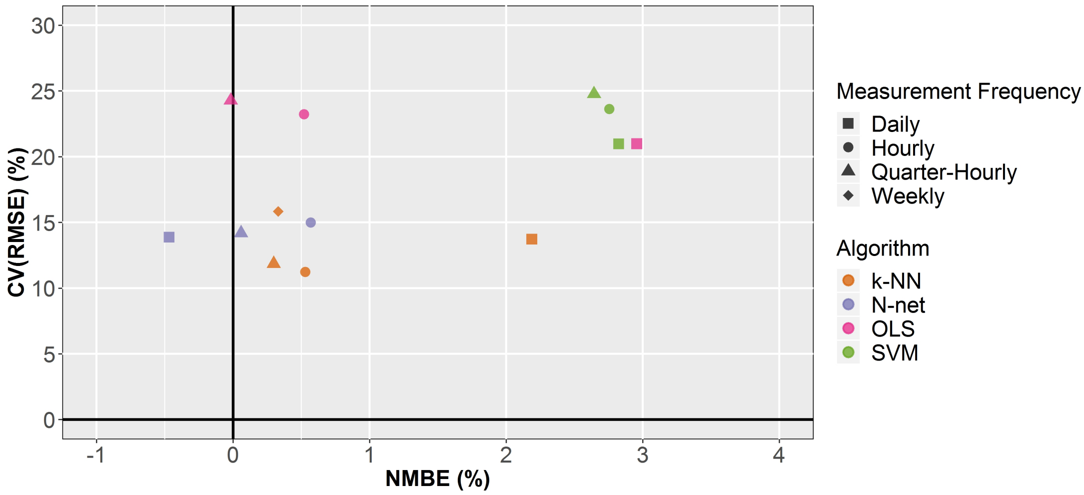


Figure 4.7: Performance of all models evaluated on testing data set

to train the hyper-parameters of each model as discussed in Section 4.4.5. This exhaustive approach led to 16 models being constructed.

In contrast to practices employed by the IPMVP and ASHRAE Guideline 14, unseen data was used to evaluate the performance of each model in the baseline period. This helps prevent over-fitting the model to the training data set, thus increasing its applicability. The SE, or RMSE, is used to calculate uncertainty in the final savings, hence, it is important that this is reliably quantified. The testing data set was normalised using the same scaling factors applied to the training data set. The performance was then evaluated by applying each model constructed to the testing data set and comparing estimated values of the dependent variable to measured values. This would not be possible without the data partition already carried out. Figure 4.7 illustrates the performance of each model in predicting the chilled water system electricity consumption during the baseline period.

The optimal model was selected based on CV(RMSE), as this metric directly impacts on energy savings uncertainty in the reporting period. A k-NN model trained using data with an hourly measurement frequency was the best performing model with a CV(RMSE) of 11.23%. A triangular kernel, 5 being the maximum number of neighbours and a distance equal to 1 were the associated model hyper-parameters.

4.5.6 Step 6 - Savings Quantification

Following the complete implementation of the ECM, the reporting period began. The information relating to data sources and characteristics documented in

Section 4.4.1 was used to gather the data necessary to quantify savings in the reporting period. It is at this point that the value of having contextualised the data comes to fruition as the data can be gathered more easily with clear semantic modelling. As with the baseline period, the raw data is gathered from both data sources and stored in a single data set.

Preprocessing consists of checking the quality of the data and aggregating it into the measurement frequency corresponding to the optimal baseline model. Quality checking is performed by assessing if the independent variables data gathered in the reporting period conforms to the range requirements of ASHRAE Guideline 14 (i.e. must not be more than 110% and no less than 90% of the corresponding baseline data). This was not an issue for this particular case study. The data was measured with a 15-minute measurement frequency and was subsequently aggregated to have an hourly measurement frequency to conform to the baseline energy model input requirements. The model was then applied to the processed data set to calculate the adjusted baseline energy consumption.

In most cases, the adjusted baseline can be directly compared to measured quantities of the dependent variable for quantification of savings. However, this is not the case in this application as there was a change in static factors during the M&V period of analysis. The construction of a production area occurred during the implementation period and was live at the beginning of the reporting period. This area houses an additional production line and has the capacity for the operation of additional production lines in the future. This increased the cooling load of the facility which requires the savings calculated to be adjusted. While the chilled water system electrical consumption increased, the cooling load also increased. Hence, had the ECM not been implemented, the load would have been satisfied in a less efficient manner with even higher chilled water system electrical consumption. This non-routine adjustment was made based on the floor area of the site increasing by 20%. This was deemed acceptable as the chilled water system services the space cooling requirements.

The energy savings in the reporting period are the difference between the measured consumption and the adjusted baseline, following the application of the non-routine adjustment. The savings were calculated to be 604,527 kWh. A mere calculation of savings without an associated level of uncertainty and confidence is of little use in ensuring reliability and completeness in M&V. The range of savings must be calculated for a given confidence interval to gain a true insight into project performance. This range is dependent on the model performance in the baseline period. The more accurate the baseline energy model, the smaller the range of savings. The IPMVP approach to uncertainty calculation detailed in Equation 4.5 was employed to calculate the range of savings in this project to be;

$$\text{Range of Savings} = 604,527 \pm (\text{t x S.E.})$$
$$= 256,485 \text{ to } 952,568 \text{ kWh @ } 68\% \text{ Confidence}$$

4.6 Discussion

4.6.1 Energy Savings in Reporting Period

The savings were calculated as being the difference between measured consumption in the reporting period and the adjusted baseline. This is illustrated in Figure 4.8. The regression model demonstrates a good capability of forecasting relative to the measured data for large parts of the reporting period, although there are periods in the summer months of 2017 in which the model actually predicts that more energy is being consumed than would have been pre-ECM. Such discrepancies are common in real world applications in which perfect ECM implementation is often not achieved.

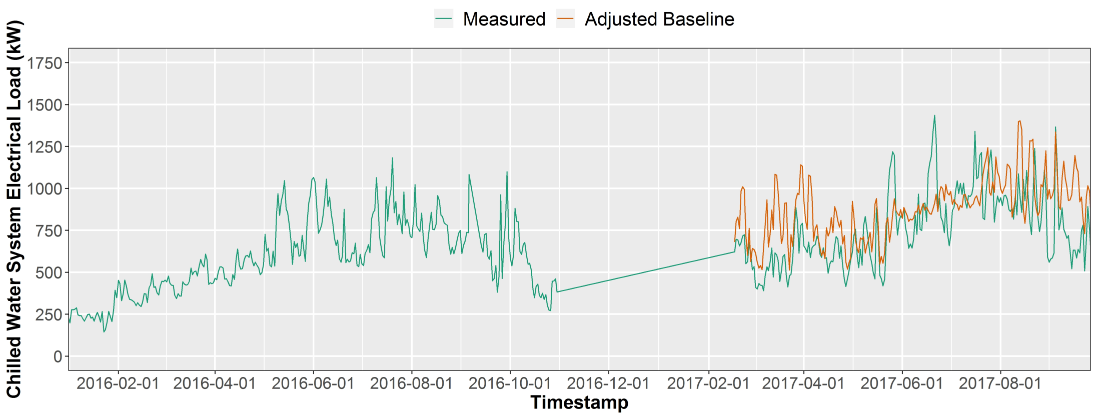


Figure 4.8: Measured consumption and adjusted baseline for entire period of analysis

4.6.2 Range of Savings

The impact of modelling error on the range of savings can be seen in Figure 4.9 with detailed results provided in Table 4.5. The results include the savings estimated by all models developed and the associated range of savings estimated with 68% confidence. The uncertainty in a project can be directly seen in the range associated with the quantified savings. The mean savings across all models is 710,558 kWh with a coefficient of variation (CV) of 39.9%. Although the savings quantified by each of the 20 models are dependent on the measurement frequency, algorithm and individual hyper-parameters, trends are evident over the set of models. The ANN regression models predict the highest savings on average in each frequency subset with an average savings of 938,927 kWh and a CV of 28.3% . The OLS models are the most consistent across all measurement frequencies with an average savings figure of 686,866 kWh and a CV of 12%, however, these savings also had the largest associated ranges. In terms of average savings, the SVM models were the most erratic with a CV of 62.6% across results.

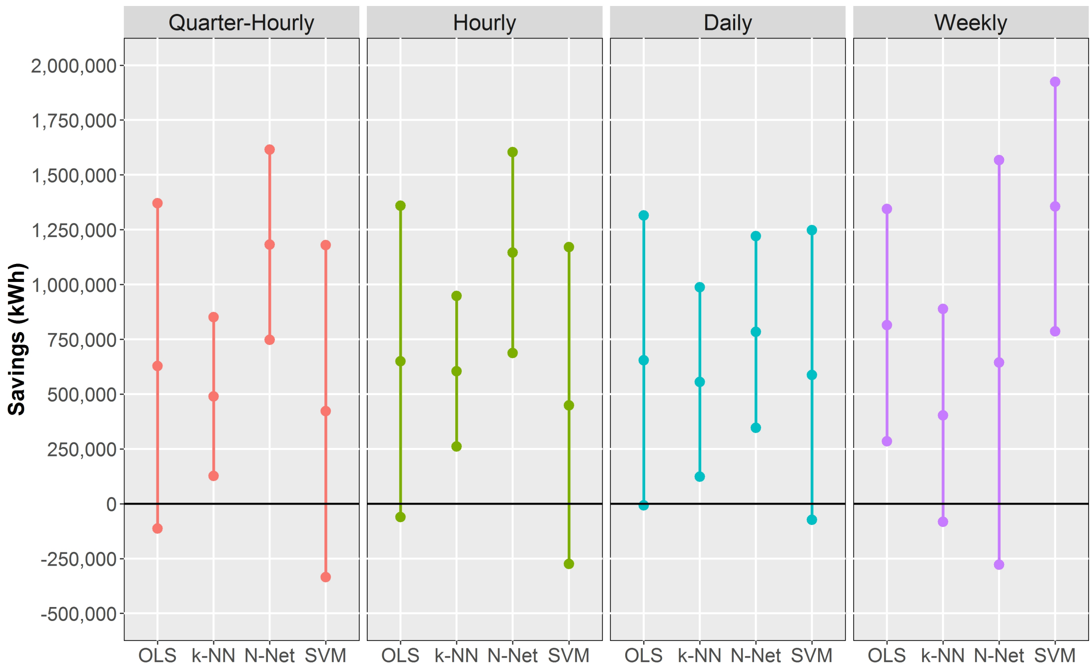


Figure 4.9: Range of savings for all models developed under varying measurement frequency with a confidence interval of 68%

This analysis shows the sensitivity of baseline energy models to algorithms, measurement frequencies and hyper-parameters. The range of savings across all 20 models is particularly distressing and this emphasises the need to maximise accuracy in baseline energy models to ensure the range of savings are small; thus, maximising confidence in the process.

Table 4.5: Savings in kWh for all models developed under varying measurement frequency with a confidence interval of 68%

Measurement Frequency	Algorithm	Minimum (kWh)	Median (kWh)	Maximum (kWh)
15-min	OLS	-113,236	628,938	1,371,112
	k-NN	126,907	489,202	851,497
	ANN	747,769	1,181,674	1,615,579
	SVM	-335,004	422,329	1,179,662
Hourly	OLS	-60,369	649,542	1,359,454
	k-NN	261,399	604,527	947,655
	ANN	687,936	1,145,835	1,603,735
	SVM	-273,881	448,371	1,170,623
Daily	OLS	-7,055	654,115	1,315,285
	k-NN	123,428	555,656	987,884
	ANN	346,580	783,606	1,220,632
	SVM	-73,504	587,279	1,248,061
Weekly	OLS	285,067	814,873	1,344,678
	k-NN	-82,744	402,815	888,374
	ANN	-278,097	644,592	1,567,281
	SVM	786,834	1,355,583	1,924,333

4.6.3 Acceptable Uncertainty

As stated in the IPMVP, uncertainty is deemed acceptable when the savings are larger than twice the standard error of the baseline energy model. For comparative purposes, each model developed was assessed to check if the uncertainty levels could be deemed acceptable. Table 4.6 contains the results of this analysis. It was found that only 3 of the 20 models developed meet the criteria defined by the IPMVP. Critically, the optimal model identified and applied for the final calculation of savings in the case study (k-NN with hourly measurement frequency) does not meet the criteria for acceptable uncertainty. As discussed in Section 4.4.5, the final model is chosen as that which has the smallest value for CV(RMSE) as this influences the final modelling uncertainty. The problem with the check carried out for acceptable uncertainty is that the models that tend to predict a higher quantity of savings are favoured, without necessarily being the most accurate models. As discussed, the ANN models predicted the highest savings on average across all measurement frequencies, while the k-NN models had to lowest CV(RMSE) for every measurement frequency. This results in the k-NN models not being deemed acceptable due to their more conservative estimation of savings. Hence, this is somewhat of a flawed process for the final evaluation of model performance.

Table 4.6: Acceptable levels of uncertainty for each model developed

Measurement Frequency	Algorithm	Savings (kWh)	Standard Error (kWh)	Acceptable
15-min	OLS	628,938	746,293	No
	k-NN	489,202	364,306	No
	ANN	1,181,674	436,313	Yes
	SVM	422,329	761,536	No
Hourly	OLS	649,542	713,806	No
	k-NN	604,527	345,010	No
	ANN	1,145,835	460,412	Yes
	SVM	448,371	726,214	No
Daily	OLS	654,115	663,477	No
	k-NN	555,656	433,737	No
	ANN	783,606	438,551	No
	SVM	587,279	663,088	No
Weekly	OLS	814,873	525,191	No
	k-NN	402,815	481,329	No
	ANN	644,592	914,651	No
	SVM	1,355,583	563,795	Yes

4.6.4 Measurement Uncertainty

The uncertainty in the savings quantified is due to the error introduced by the baseline energy model. Measurement error was omitted from this analysis to analyse the performance of the regression model in isolation. It is critical that the range of savings reported is minimised to maximise confidence in the M&V process. It is important to note that the measurement error is likely to increase as more independent variables are required to model the baseline energy consumption. In particular, when these measurements are recorded using metering infrastructure on-site, each will have an associated uncertainty. This is in contrast to the more simplistic modelling solutions that rely on weather and occupancy data.

4.7 Conclusions

The research presented in this chapter offers a novel approach to utilise machine learning techniques for energy savings verification. A focus is placed on minimising uncertainty introduced by the energy model in the quantified savings. A definitive methodology was developed to provide explicit guidance on the application of machine learning for the purposes of maximising the accuracy with which M&V can be carried out. M&V practitioners do not need to have knowledge of the inner working of the modelling algorithms, as the step-by-step approach to the problem includes performance checks that must be met to ensure accuracy. A whole-facility approach is adopted to widen the scope of analysis, while isolating the energy system in which the ECM is implemented. This blended approach defines a novel boundary of analysis and makes use of all data recorded across the facility, without requiring the energy savings to be large relative to the site load (IPMVP Option C constraint). The methodology has been designed to be robust enough to be compatible across the spectrum of M&V projects.

The proposed methodology is of particular benefit in circumstances with limited metering infrastructure directly related to the energy system under analysis. Evidence of which can be seen in the case study where M&V would not have been possible without the installation of additional metering equipment and a data gathering period that would delay the ECM implementation. The methodology was directly applied using real-world data for a large biomedical manufacturing facility. A total of 20 models were developed in the baseline period using an exhaustive approach. The optimal model was identified as a k-NN regression model trained with data measured hourly. The CV(RMSE) of this model was 11.23%, while the $R^2_{adjusted}$ was calculated as being 0.93. The reporting period was 222 days in duration with estimated energy savings over this period being 604,527 kWh. Critically, following the quantification of the associated uncertainty, these savings were found to range from 256,485 to 952,568 kWh at 68% confidence. The range of savings for all 20 models constructed were investigated to show the impact model performance has on final savings. This highlighted that the procedure used to assess acceptable uncertainty favours models that estimate higher savings in the reporting period, rather than those that perform better in the independent cross validation testing.

It is important to note that the conditions in the case study are representative of the issues that exist across the industrial buildings sector. These include a lack of sufficient metering, poor data quality and changing static factors. The application of the methodology demonstrates its ability to overcome these issues and quantify energy savings with an acceptable level of uncertainty. The level of uncertainty achievable is dependent on individual project characteristics. Specifically, the relationship between the dependent variable and available independent variables is the limiting factor in minimising uncertainty. The proposed methodology seeks to realise the achievable uncertainty by utilising the available resources. In the case study, the industrial building did not represent ideal conditions and the results illustrate this. The relationship between the chilled water electrical load and the 10 independent variables had an $R^2_{adjusted}$ value of 0.663 using an OLS regression model. This was for a 15-minute measurement frequency which represented the poorest value of all the frequencies analysed. As this is on the lower end of statistically significant values, the achievable model performance is limited by this relationship. Evidence of this can be seen in the large range of savings in the final results. That said, the problematic nature of the case study enabled a robust assessment of the methodology to take place.

The approach taken by the methodology reduces the need to install additional metering infrastructure. This has the potential to greatly reduce the resources required to complete accurate M&V in any given project. Machine learning techniques are able to extract relevant knowledge from the available data set and utilise it to develop the baseline energy model.

Chapter 5

An Intelligent Framework for Integration with M&T

5.1 Introduction

5.1.1 Overview

The baseline energy modelling methodology presented in the previous chapter empowers M&V practitioners in the use of machine learning techniques to maximise accuracy in performance verification. These techniques are essential tools in the transition to M&V 2.0 practices in which automated analytics are applied to large energy data sets. Prior to the development of the modelling methodology, the value of these approaches was demonstrated in Chapter 3. This constitutes the first step in evolving practices to the desired more advanced and useful state, as it populates a significant knowledge gap in the sector.

Following the development of the prescriptive baseline energy modelling methodology, there remains a deficit of guidance on the application of such practices in an M&V 2.0 manner. How can the modelling methodology be applied in an automated fashion and in near real-time to provide ongoing quantification of savings? As discussed in Chapter 2 (Section 2.6), the M&V 2.0 solutions available on the market today are dominated by the commercial buildings sector. In addition to this, the established protocols, such as the IPMVP and ASHRAE Guideline 14, have not been developed with automated analytics and granular data sets in mind. Thus, the sector is devoid of any guidance on the successful implementation of M&V 2.0 practices. Additionally, how can these practices be integrated into ongoing energy management tasks?

This chapter presents a novel framework that enables automated, real-time performance verification of ECMs in industrial facilities. The framework utilises the novel baseline energy modelling methodology developed to harness the power of available data. The optimal baseline energy model identified is continually applied in an automated manner for the duration of the reporting period to enable real-time quantification of savings. A performance deviation detection check is also presented with a view to enabling exception reporting. This ensures the early identification of performance degradation, which is critical to maximising savings realised. The benefits of the proposed M&V 2.0 framework are demonstrated using a case study in a large-scale manufacturing facility.

5.1.2 Background

As introduced in Chapter 2 (Section 2.6), the field of performance verification in energy systems is evolving with automated, advanced analytics becoming increasingly common. The term M&V 2.0 is being used to identify approaches to M&V that employ these more powerful and less labour intensive techniques. This represents a shift from the traditional, retrospective and static approaches to modern, real-time and dynamic processes. The use of granular data coupled with automated processing has been identified as a necessary development to progress the application of M&V as a vital tool in the transition towards low-carbon economies (Franconi et al. 2017).

Although it is not often discussed when reviewing this movement, the transition to more mature practices offers the opportunity to finally manage the integration of M&V into ongoing energy management tasks. It is proposed in this thesis that the area of M&T is the most suitable element of EnMSs to accommodate this integration. Success in this regard will not only ensure M&V is a more dynamic process, but it will also allow for it to become reactive through the use of AFDD techniques. In a comprehensive review of these methods, it was found that AFDD tools are no longer seen as standalone solutions. They are being integrated with existing tools and practices to maximise the efficiency with which systems operate (Bruton et al. 2014). Thus, it is sensible to apply this same logic in the field of performance verification.

The necessity for energy savings to persist for the duration of an ECM's lifetime has been discussed at length. One such example of persistence presenting challenges is the energy efficiency obligation scheme in operation in Ireland pursuant to the EU's EED. This scheme places the responsibility on energy suppliers to achieve savings relative to their market share. In response to this, energy suppliers offer financial incentives to industrial customers to achieve energy savings. This system results in payments being exchanged for energy savings with an approved M&V report. It is important to note that the duration of the reporting period in these projects does not exceed 12 months, in general. Therefore, there is no guarantee that savings will persist until 2020 without revisiting the initial M&V. In this chapter, a novel, robust and open-source framework is developed that not only seeks to satisfy M&V 2.0 needs, but also bridge the gap between M&V and M&T to ensure long-term persistence of savings. As stated in Chapter 2 (Section 2.6), it is widely accepted that M&V 2.0 is defined as applying automated analytics to granular data sets to provide ongoing, near real-time performance verification (Franconi et al. 2017). Despite this, there is little clarity over the exact conditions which satisfy this definition. Thus, the characteristics of M&V 2.0 can be broadly defined as follows:

- **Granular Data Sets:** The advent of digitalisation has resulted in AMI being common place in modern industrial facilities. This has enabled greater insight into energy consumption behaviour as the quantity of measured data available increases. High resolution (hourly or more granular) data is now available from main site meters to extensive sub-metering systems. The availability of such data allows for more granular analytical approaches to be applied.
- **Automated Analytics:** The M&V process can be streamlined using advanced analytics implemented using software solutions. Ongoing monitoring and quantification of energy savings in near real-time is made possible using such approaches.

5.2 Research Questions

Solutions to the following research questions were desired, thus they were used to lead the development of the M&V 2.0 framework outlined:

1. How can machine learning techniques be implemented in M&V 2.0 applications?
2. Can advanced analytics and granular energy data be utilised to increase the reliability and usefulness of M&V by providing near real-time quantification of savings?

3. Can a means of exceptional reporting be incorporated into the process to automatically identify deviations from expected performance?

4. Will a smooth transition from M&V to M&T ensure energy savings persist over a projects lifetime?

In addition to these chapter specific research questions, RO4 is also addressed by the work detailed.

5.3 Methodology

In keeping with the design philosophy of the baseline energy modelling methodology, the proposed M&V 2.0 framework seeks to take advantage of the data recorded by the omnipresent AMI in modern industrial facilities. AMI is now common place in most facilities operating an ISO 50001 certified energy management system. It is critical that the resources required to carry out M&V are not increased when employing these large quantities of data. The automated nature of this framework aims to avoid this risk. The framework is built around the baseline energy modelling methodology detailed in Chapter 4, while the features added allow the models developed to be deployed effectively on an ongoing basis.

The framework is sub-divided based on the period of analysis at each stage of the project. These are the baseline, implementation, reporting and persistence periods. The persistence period is a new period of analysis that has been added to the performance verification process and it occurs following the completion of traditional M&V. It is designed to enable a smooth transition from M&V to M&T and subsequently enabling savings to be maximised over an ECM's lifetime.This is in contrast to quantifying savings over the reporting period in isolation. Figure 5.1 provides a graphical illustration of the framework developed.

5.3.1 Baseline Period

Data Gathering

The two primary resources required for accurate M&V are skilled practitioners and sufficient data that accurately accounts for the energy system's state. An approach that can utilise available data and automatically compute savings in complex environments is essential to minimising the overall costs of completing M&V. An evaluation of the available data must be completed to assess the ability of this data to be used for reliable performance verification. If the data available

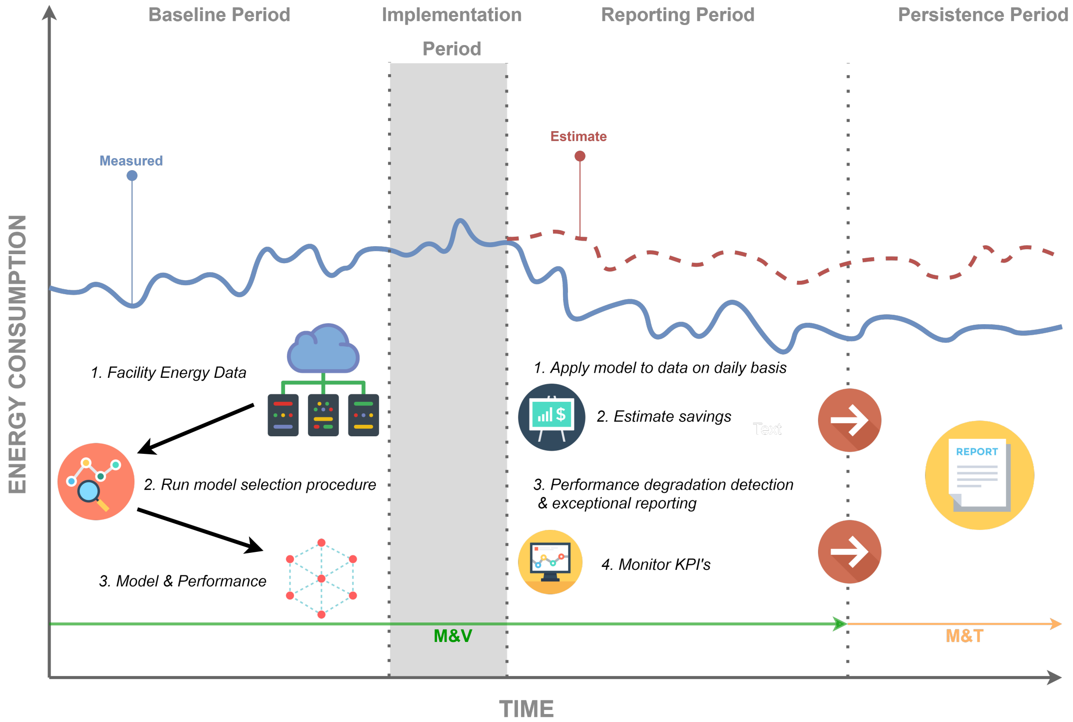


Figure 5.1: Illustration of the M&V 2.0 guidance framework developed

is insufficient, additional metering infrastructure must be installed, thus increasing project costs. This can also delay project implementation as baseline period data must often be gathered with the new metering equipment. This stage of the process consists of identifying suitable data sources and recording the characteristics of each relevant data source. This should include the type of data, measurement frequency, storage methods and access protocol. The objective is to outline a means of accessing data from each distributed data source to enable data extraction. This is a step that is covered comprehensively in Chapter 4 (Section 4.4.2).

Baseline Energy Model Development

The construction of an accurate model of the energy system's performance in the baseline period is a critical step in the M&V process. An accurate baseline energy model can then be applied post-ECM implementation to normalise post-ECM consumption to pre-ECM conditions; a requirement for the computation of final savings. Modelling error is a prominent source of uncertainty in M&V and thus, it must be kept to a minimum to ensure the uncertainty associated with the final savings is within acceptable limits.

It has been discussed in Section 5.1.2 that a lack of prescriptive guidance on the construction of a baseline energy model in the most commonly used protocol is a hindrance to their effective implementation. The baseline energy modelling methodology developed in the previous chapter employs machine learning techniques to populate this knowledge gap in the field. Crucially, the methodology provides step-by-step guidance on the application of advanced regression algorithms to construct the optimal baseline energy model for any given project. This methodology has been incorporated into the proposed framework as it is technology agnostic and enables the use of efficient analysis of large data sets. It is at the discretion of the individual M&V practitioner as to which modelling algorithm to employ, although the exhaustive algorithm, which utilises four machine learning regression techniques, detailed already in Chapter 4 is recommended.

A key feature of the modelling methodology is a computationally efficient wrapper-based feature selection algorithm that can be employed to automatically identify relevant independent variables for use in model construction. This is significant in reducing the need for subject-matter knowledge on each individual ECM, as it is one of the objectives of feature selection to provide a better understanding of the underlying process that generated the data (Guyon & Elisseeff 2003). This is achieved by identifying correlations between complex variables in large data sets.

The advanced regression techniques applied by the model are multiple OLS, k-NN, multi-later perceptron feed-forward ANN and SVM. An exhaustive approach to modelling is employed in which each algorithm and a range of measurement frequencies are utilised to produce an array of baseline models.

Identification of Optimal Model

The successful application of the modelling methodology results in the construction of a number of baseline energy models. This exhaustive approach to energy modelling requires the optimal model to be identified. The performance of each model is evaluated on a previously unseen data set. This is achieved by partitioning the data available in the baseline period into two data sets. An 80:20 random split ratio is used to generate training and testing data sets. This enables models to be constructed using 80% of the baseline data and their performance evaluated on the remaining 20%. The uncertainty is calculated using the Equation 4.5 defined in Chapter 4.

The output from the baseline period analysis is the optimal model of the energy system's performance prior to the implementation of the ECM and its associated uncertainty. This model can them be deployed throughout the reporting and persistence periods in an automated fashion.

5.3.2 Implementation Period

It is important to clearly define the implementation period in any M&V project. This period is used to fully implement and commission all ECMs. A poorly defined implementation period could result in irrelevant data being used for model construction and/or deployment. No analysis is carried out during this period as the energy system is in transition.

5.3.3 Reporting Period

Model Application

The model produced in the analysis completed prior to the implementation of the ECM is to be applied at regular intervals in the reporting period. The frequency with which the model is applied is dependent on the measurement frequency of the data used to train it. Specific to each individual project, the model is to be applied at the same frequency with which the training data in the baseline period was recorded. For example, a model constructed using data with a 15-minute measurement frequency should be applied every 15-minutes to new data gathered for the model features. This enables real-time quantification of savings. Barriers to this application are generally related to data availability. It is possible for data to be recorded in 15-minute intervals, but only made available to the end-user every 24 hours due to the configuration of the data pipeline. In all cases, the highest possible frequency of application should be employed.

Real-time Savings Quantification

The energy savings are calculated using the IPMVP approach defined in Equation 5.1. This is a measure of the success of an ECM implementation and the continued operation of the system. Non-routine adjustments are project specific measures taken to adjust the reporting period conditions. They are necessary when static factors change over the project lifetime. For example, changes in the size of a facility or manufacturing process schedules would require a non-routine adjustment as the baseline energy model was constructed under different operating conditions.

$$
\begin{aligned}
\text{Savings} = {} & \text{Predicted Consumption for Reporting Period Using Baseline Model} \\
& - \text{Reporting Period Measured Consumption} \\
& \pm \text{Non-Routine Adjustments}
\end{aligned}
\tag{5.1}
$$

Exceptional Reporting of Performance Deviation

Prior to the implementation of an ECM, a feasibility study will generally be carried out to assess the potential savings and associated costs. This will result in an estimation of performance. If this has not taken place, an engineering first-principles approach should be used to estimate the savings that will be achieved. This estimation of savings can be compared with the actual system performance to set upper and lower control limits to identify performance deviations.

Energy performance contracts (EPCs) offer a more rigid savings estimations that can be employed. An EPC is a finance mechanism used in the energy services industry in which customers 'pay for performance'. In cases where EPCs are in place, then this figure should be used as the primary estimation of savings.

The actual performance found using the baseline energy model is compared with the expected performance to establish if the savings are on track. Any deviations from expected performance triggers an exception report to the engineering team. As a rule of thumb, a 20% deviation is defined as a deviation from expected performance. This threshold was arrived at after considering the potential error in the preliminary estimation of savings used to compute it. Practitioners may chose to employ a lower threshold for stricter control. This automated system provides an insight into system performance, enabling corrective action to be taken to maximise the savings realised.

Monitor KPIs

As suggested by ASHRAE in Guideline 14, the model can only be applied for periods where independent variables are no more than 110% of the maximum and no less than 90% of the minimum values of the same variables used for constructing the baseline energy model. This is a straightforward step that can easily be automated. If independent variables stray outside of these bounds, then the error metrics associated with them are no longer valid. The model must be retrained with more suitable variables in these circumstances.

5.3.4 Persistence Period

The persistence period occurs outside the scope of traditional M&V. This is the point at which M&T takes over the evaluation of system performance. This new period of analysis enables performance evaluation to be an ongoing task.

Persistence Plan

A plan is required to ensure persistence of savings over the lifetime of an ECM. This includes the continuous operation of the automated system for performance tracking. The persistence plan should also detail responsible individuals in cases where the performance tracking system must be revisited, such as independent variables no longer being within bounds. Integrating these key elements of M&V into the M&T process allows for longer term savings tracking.

Adjustments

Adjustments are required on a project-by-project basis. This includes reconstruction of a baseline model in cases where independent variables are no longer relevant and applying scaling factors when significant changes occur to the facilities' operating conditions.

5.4 Case Study: Results and Discussion

The proposed framework was applied to quantify the savings resulting from an ECM carried out on a set of AHUs in a large biomedical manufacturing facility in Limerick, Ireland. The facility operates a continuous production process on a 24/7 basis. The ECM consisted of optimising the control logic for each individual AHU. The new control logic is more intelligent than the previous one, with an ability to respond to the space heating and cooling requirements of the areas served. This is in contrast to the static system in place pre-ECM, which supplied a fixed volume of air to each area. The logic utilises variable speed drives (VSDs) already in place to vary the volume of air supplied depending on the requirements of the environment being treated. In existence pre-ECM, were electricity consumption meters on each AHU. Therefore, the decision was made to assess the savings in the total electricity consumptions of all AHUs, i.e. the cumulative consumption of all individual units.

As with the two preceding case studies presented in this thesis, the framework was applied using the open-source programming language R. The automated application of the model and checking of KPI's in the reporting period were achieved using a scheduled task. Specifically for this case study, the software utility cron was used on a Linux server.

5.4.1 Baseline Period

The modelling methodology discussed in Section 5.3 was applied to identify the independent variables relevant to the total AHU electricity consumption and subsequently, model the performance of the system in the baseline period. The baseline period was selected to begin on January 1st, 2016 and it concluded on October 4th, 2017. This period encompassed more almost two full 12-month cycles of analysis of the system, hence covering a wide spectrum of operating conditions. This is important in ensuring the model's validity is maintained in the long-term.

The optimal approach that minimised model uncertainty was a k-NN model trained with data having an hourly measurement frequency. 18 independent variables from across the site were used to construct this model. Figure 5.2 contains box and whiskers plots of each feature to summarise the spread of values. These were selected based on statistical significance to the dependent variable. All data was gathered using existing metering infrastructure.

The performance of this optimal model was quantified as having a standard error of 15.99 kW when evaluated on the unseen testing data set. Figure 5.3 illustrates the fit of the model on a sample of data in the baseline period.

5.4.2 Implementation Period

The implementation period began on October 5th, 2017 and concluded on November 11th, 2017. No installation or commissioning works were carried out outside of this period.

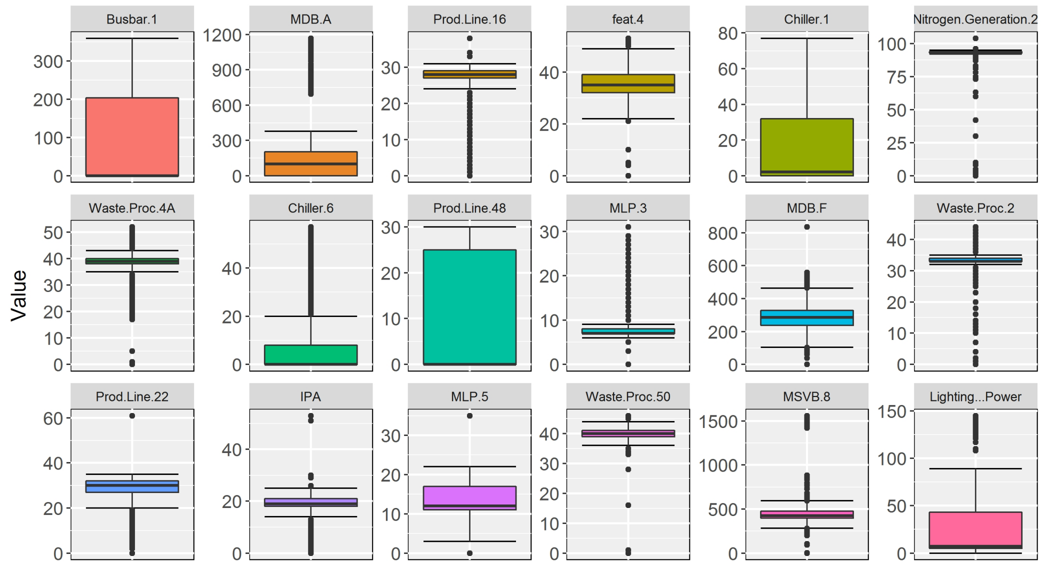


Figure 5.2: Box and whisker plots of each independent variables used to construct baseline energy model.

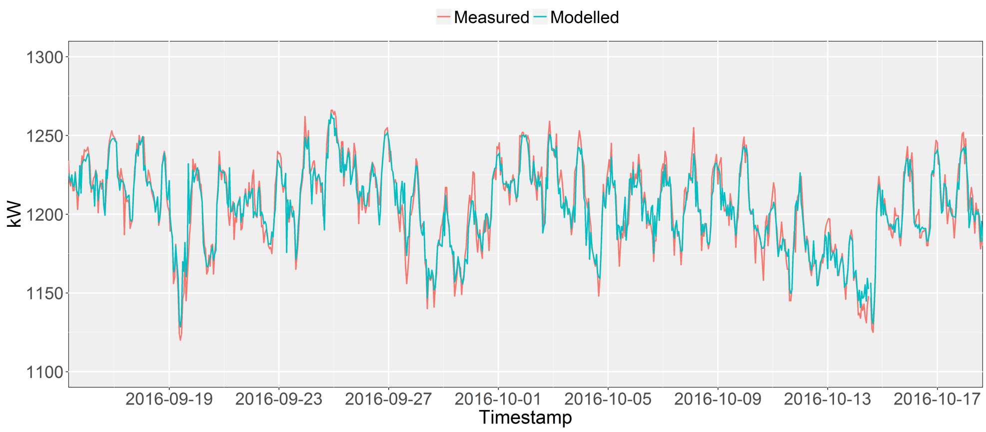


Figure 5.3: Sample of model fit in baseline period.

5.4.3 Reporting Period

The reporting period began on November 12th, 2017 and concluded on December 1st, 2017. This is a relatively short period of analysis that was limited by research project parameters. Despite this, the lack of seasonality in the electricity consumption of the AHUs and the wide array of conditions in the baseline period ensure the findings are reliable, while providing an insight into the benefits of long-term application.

The optimal model was trained using hourly data, hence this is the minimum frequency with which it could be deployed. The application of the optimal model over the 19.5 days reporting period resulted in energy savings being quantified to be 177,962 +/- 12,334 kWh with a 90% confidence interval. This is equivalent to a 380.1 kW reduction in electrical load on the system. The uncertainty associated with the savings is just 8.6% of the allowable uncertainty as defined by the IPMVP.

A simulation approach was used to demonstrate an implementation of the performance deviation detection (PDD) system. This consisted of simulating two periods in which system performance deviated from expected levels and assuming a figure for estimated savings. This approach was required as this information was not yet available at this time. Hence, it was assumed that a feasibility study carried out prior to any implementation works being carried out estimated a reduction in the electrical load of the AHUs of 385 kW. This estimation was based on assumed VSD motor efficiencies and run-hours and perfect implementation for the duration of the reporting period. This figure was used to develop a rule that could be employed to identify periods of performance degradation. If the actual savings found using Equation 5.1 were less than 80% of the expected savings (i.e. 385 kW) for 4 consecutive hours or more, then an exception report is generated. This alerts the on-site facilities team to investigate and take the necessary corrective action. A graphical representation of two periods of performance degradation identified is included in Figure 5.4. Corrective action was taken to ensure performance returns to expected levels. Thus, the savings realised can be maximised. This would not be possible using a traditional M&V approach as the savings are not quantified until the reporting period is concluded and as one of the degradation events occurred within 3 days of implementation, it is unlikely that corrective action would be taken in sufficient time.

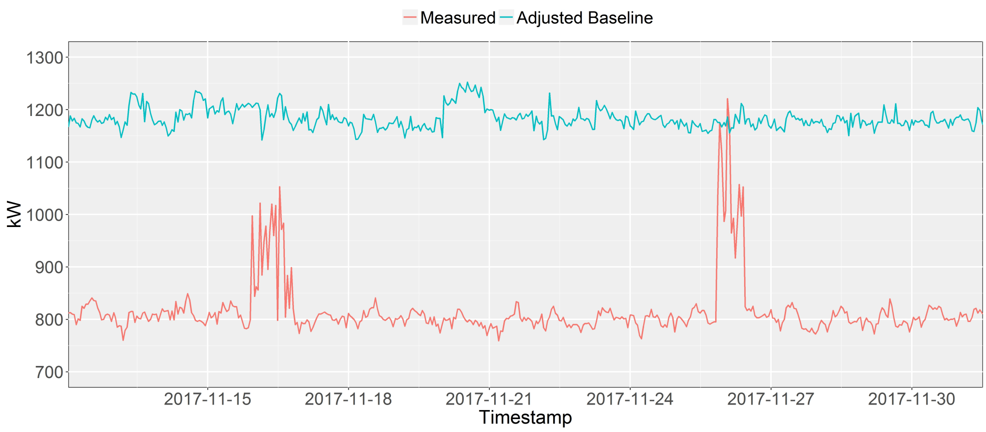


Figure 5.4: Illustration of performance deviations, associated alerts and corrective actions.

5.4.4 Persistence Period

A persistence plan was developed to ensure the maximum possible savings are realised over the lifetime of ECM and that the ongoing monitoring of savings is integrated into regular M&T activities. This plan was agreed with the on-site facilities team to ensure responsible parties are identified for possible future works. Additionally, the action to be taken should the operating conditions of the site significantly change is outlined in the persistence plan.

5.5 Conclusions

Accurate, reliable and efficient M&V of energy savings is a necessary tool in tracking the performance of energy projects. To continue to play an effective role in future energy systems, M&V 2.0 must become common place across residential, commercial and industrial applications. A technology agnostic framework for automated, real-time M&V was developed to offer a solution to this challenge and to populate the knowledge gap that currently exists in the sector. This is a useful tool that can be employed to ensure M&V evolves to a more mature state of operation. The benefits of the proposed framework were demonstrated using a case study. Two instances of performance degradation were automatically identified, allowing corrective action be taken.

The proposed approach represents an evolution from static retrospective M&V to more powerful, efficient and dynamic M&V solutions. This simple means of performance degradation identification is incorporated to enable a smooth transition from short-term M&V to long-term M&T, thus ensuring savings persist over a projects lifetime. Additionally, the framework will have an increased applicability going forward as stronger energy policy measures take hold and ISO 50001:2018 places a renewed emphasis on demonstrating clear energy performance improvements.

The framework detailed offers a resource for M&V practitioners to utilise in adopting M&V 2.0 practices. The technology agnostic nature of the framework ensures applicability across the broad spectrum of ECMs. However, the framework was developed from a theoretical perspective and like the established M&V guidance documentation, the desire to maintain wide ranging applicability can lead to gaps in knowledge when applying such approaches in the real-world. Therefore, an implementation of the M&V 2.0 guidance framework would further aid practitioners in its application, while also demonstrating a proof of concept. Furthermore, the case study presented in this chapter could be advanced through the gathering of additional post-ECM data.

Chapter 6

IntelliMaV: A Cloud Computing M&V 2.0 Application

6.1 Introduction

6.1.1 Overview

The framework developed in the previous chapter provides guidance on the use of data-driven machine learning techniques in M&V 2.0 applications. This aids the adoption of these more advanced and powerful practices that in turn enable near real-time quantification of energy savings with a higher degree of accuracy. This framework builds on earlier analysis presented in Chapter 3 which highlights the potential performance improvements that can be achieved through advancing the energy modelling process. Subsequent to this, a baseline energy modelling methodology was developed to provide prescriptive guidance to M&V practitioners. The M&V 2.0 framework provides all-encompassing guidance for the successful delivery of automated and advanced analytics on large data sets with performance deviation detection.

Given that the void in guidance documentation has been populated in the previous chapters, the primary barrier to the adoption of M&V 2.0 practices are the tools available to practitioners. This chapter aims to populate the knowledge gap in the industrial buildings sector by presenting a novel cloud computing-based application, IntelliMaV, that applies advanced machine learning techniques on large data sets to automatically verify the performance of ECMs in near real-time. Additionally, a performance deviation detection system is incorporated, ensuring persistence of savings beyond the typical period of analysis in M&V.

This system represents a proof of concept implementation of the M&V 2.0 framework presented previously.

IntelliMaV allows M&V practitioners to quantify energy savings with minimum levels of uncertainty by applying powerful analytics to data readily available in industrial facilities. The use of a cloud computing-based architecture reduces the resources required on-site and decreases the time required to train the baseline energy model through the use of parallel processing. The robust nature of the application ensures it is applicable across the broad spectrum of ECMs in the industrial buildings sector. A case study carried out in a large biomedical manufacturing facility demonstrates the ease of use and benefits realised through its adoption.

6.1.2 Background

As has been discussed, energy efficiency is a potent resource capable of delivering demand-side energy savings; however, a challenge lies in the processes used to account for savings resulting from energy efficiency measures. The barriers to investment in cost effective ECMs were reviewed in depth in Chapter 2. The findings of this review showed that hidden costs and staff constraints are two primary factors that contribute to the existence of the energy efficiency gap (Thollander & Ottosson 2008, Meath et al. 2016). Hidden costs in energy efficiency include the often overlooked costs of performing performance verification. As stated earlier, typical M&V costs are 1-5% of total project costs when the IPMVP Option A is employed and 3-10% using Option B. It is important to note that the assumptions employed in Option A contribute significantly to the uncertainty in final savings. Thus, while costs are reduced through the use of this approach, the reliability of savings quantified is also impacted on. In any case, the minimisation of these costs is critical to maximising investment in energy efficiency as a resource in the transition towards low-carbon economies. Automating the performance verification process presents an opportunity to deliver on this objective and it has been identified as a key characteristic of M&V 2.0 practices (Franconi et al. 2017).

Automating the M&V process offers multiple benefits to the industry beyond cost reduction. One such benefit is in removing the current barrier that exists that is a lack of skills in the workforce preventing the implementation of energy efficiency measures. This barrier will only become more prominent as practices evolve and advanced analytics become more common place in the industry. Hence, a solution is required that enables practitioners that do not posses expert knowledge of these advanced data-driven techniques to successfully implement M&V 2.0 methods in a supervised manner. Despite the presence of a number of solutions for M&V 2.0 in the industrial buildings sector (reviewed in Chapter 2, Section 2.6), there is no one solution that enables practitioners to apply advanced analytics on large data sets without possessing knowledge of the underlying algorithms. The application presented in this chapter automates the critical steps of the process that require this knowledge including data cleaning, feature selection, model application, model evaluation and deployment.

Additionally, it has been identified that few tools are capable of performing remote M&V without the need for locally installed software (Kupser et al. 2016). Hence, a novel software architecture is required to offer a solution to this void in the industry. In this chapter, a novel cloud computing application that utilises machine learning techniques to automatically quantify energy savings in near real-time is presented in detail. This enables the M&V practitioner to apply powerful analytics without knowledge of the underlying algorithms. The cloud computing-based infrastructure of IntelliMaV minimises the computing resources required on-site for implementation. Standardisation in each step of the process ensures reliability and trust in the results output by this black-box approach.

6.2 Research questions

The findings of an extensive review of published literature in the research area (detailed in Chapter 2) aided the formulation of the following research questions, which were used to lead the software development and analysis detailed in this paper;

1. Can a sufficiently robust M&V 2.0 software solution be developed to populate the knowledge gap in the industrial buildings sector?
2. Is it possible to perform near real-time M&V without increasing the cost of M&V?
3. How can the power of existing M&T infrastructure be utilised to minimise the resources required for performance verification?
4. Can a PDD system be integrated into the M&V process to ensure long-term persistence of energy savings?

In addition to these chapter specific research questions, RO5 is the overarching motivator for the works detailed.

6.3 Methodology

As outlined in Section 6.1.2, the research presented in this chapter offers a novel M&V 2.0 solution that automatically applies data-driven, machine learning algorithms for accurate, reliable and robust energy savings quantification. The advantages of the proposed solution include the following three key features;

1. IntelliMaV is tailored for the under-represented industrial buildings sector and the vast quantities of energy and operational data available in modern facilities.

2. The development of the baseline energy model is performed with ease in an automated yet supervised fashion. This enables the practitioner to utilise the already proven accurate machine learning techniques without expert knowledge of the underlying algorithms. The cloud computing-based system architecture is critical to delivering this software as an efficient and mobile service.

3. A PDD system, based on expected energy savings, has been integrated into the M&V process to ensure savings persist over the entirety of a projects lifetime. This offers a unique solution to the ongoing challenges faced in adopting M&V as an on-going energy management task.

6.3.1 Application Architecture

The system architecture of IntelliMaV is illustrated in Figure 6.1 with three clearly defined tiers; the user, cloud and site. This configuration was utilised to minimise the computing resources required by the M&V practitioner to accurately verify performance of an ECM. Additionally, this architecture enables parallel processing of model training algorithms, thus reducing the time required to complete the process. An overview of each computing tier is included below.

- **User tier:** The practitioner (or user) accesses IntelliMaV via a web-browser and uses this interface to oversee and manage the entire process. The user interface is split into three segments; model training, model deployment and the savings tracker dashboard. All segments are protected by a login portal. The user is required to input information on the specific ECM being evaluated. This is an interactive process with the user receiving feedback at each stage. This is important in ensuring transparency throughout the project. Figure 6.2 shows an example of the interactive nature of the application.

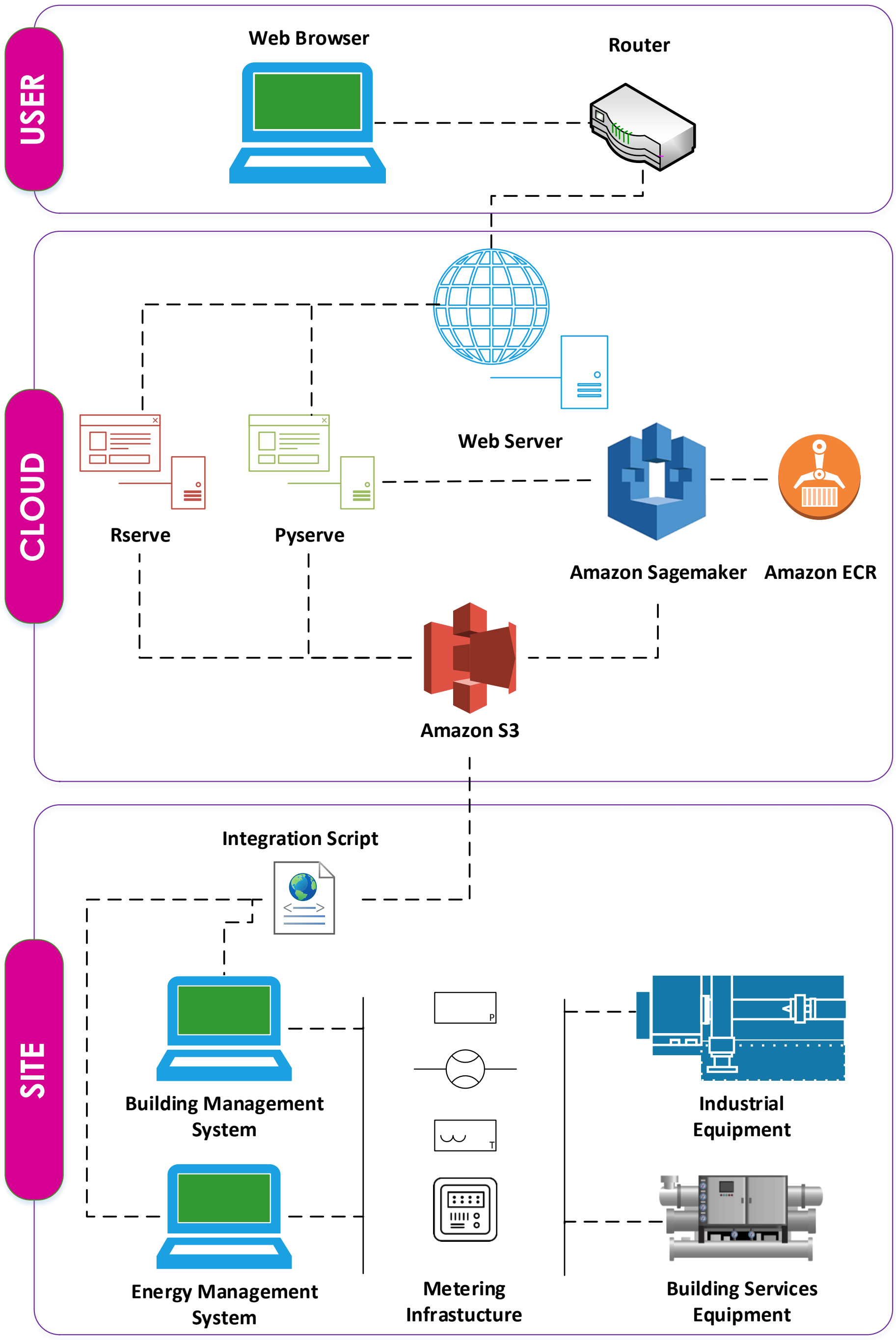


Figure 6.1: High-level M&V 2.0 application architecture

In this case, once the user understands the process outlined and clicks to apply the feature selection algorithm, the page dynamically updates and returns the result of the algorithm; the feature set. Further examples of the graphical user interface (GUI) are included in Appendix A.

- **Cloud tier**: A Virtual Private Cloud (VPC) hosts the cloud computing infrastructure required to run the application. A web server is exposed to the public and manages all incoming and outgoing requests. The analytics are handled by the application server with Amazon Web Services (AWS) employed to maximise the efficiency of tasks. The application server hosts two virtual servers, local to the VPC, running in parallel. One of these servers, referred to as '*Rserve*', executes code written in the R programming language and is executable via a Representational State Transfer (REST) application programming interface (API) exposed to the VPC. This is used for all data retrieval, cleaning, feature selection and model deployment. The second server, known as '*Pyserve*', executes python functions via a REST API for the purposes of model training. This server is used to communicate directly with AWS. A range of AWS products are employed by IntelliMaV. AWS Simple Storage Service (S3) provides object storage through a web service interface, while AWS Sagemaker is a fully-managed platform for the construction, training and deployment of machine learning models in a server-less environment. Finally, AWS Elastic Container Registry (ECR) is a fully-managed Docker container register used for storing, managing and deploy Docker container images.

- **Site tier:** The infrastructure on the industrial site is connected to the cloud computing tier via a data pipeline. It is important to note that a number of different data pipelines can be implemented to get the raw input data to a remote storage service. A single solution has not been designed as cross site applicability is required. Data recording and storage mechanisms vary amongst industrial sites, hence, it is at the discretion of the individual user as to which data pipeline is implemented. Chapter 2 (Section 2.5.4) provides an overview of some modern solutions available.

6.3.2 Model Training

The pre-ECM process consists of five stages; project description, data input, feature selection, data cleaning and model construction. Each step has been covered comprehensively in the modelling methodology developed in Chapter 4 and M&V 2.0 framework presented in Chapter 5. Therefore, only a description of the specific implementation in this application is detailed in this section. The objective of each stage and tasks involved are discussed below.

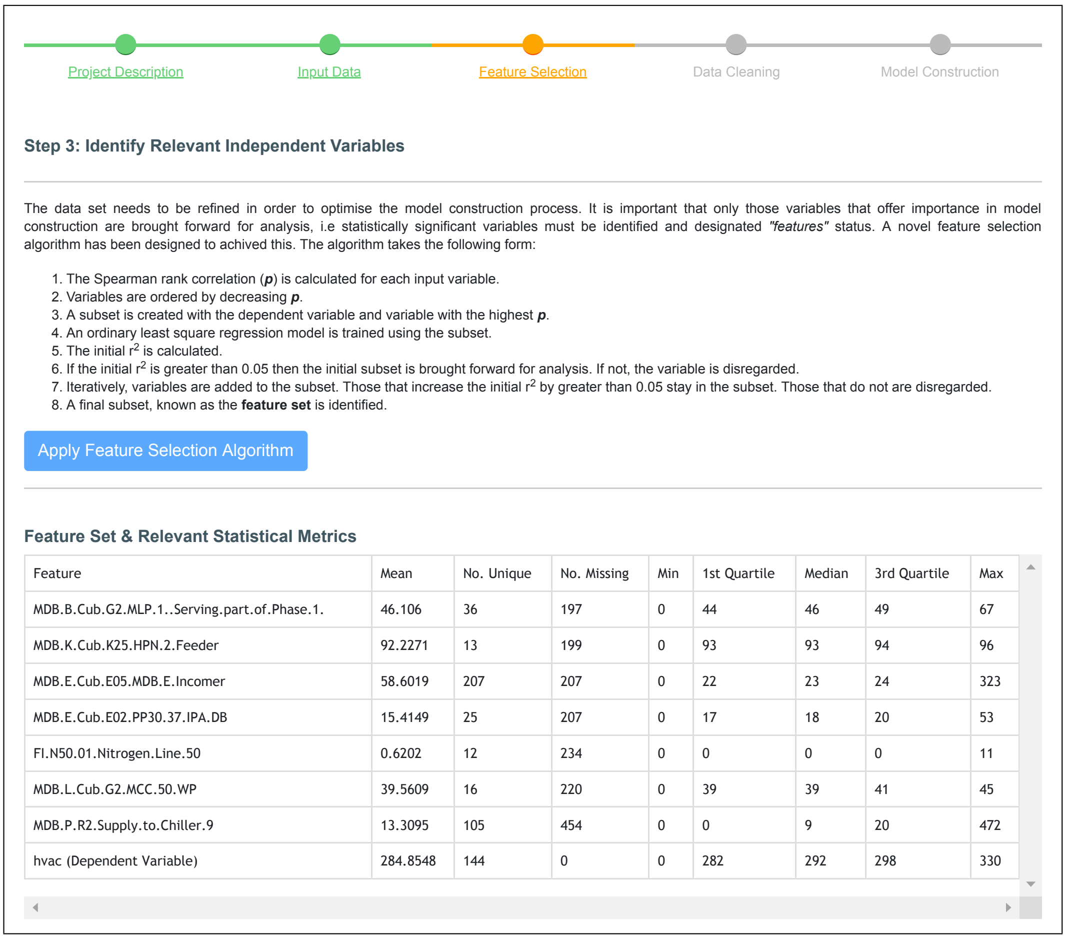


Figure 6.2: Example of IntelliMaV user interface

Project Description

The user is required to input project specific information that is known pre-ECM. This includes the project title, contact email and a description of the ECM to be implemented. This information is used at a later stage to populate the M&V report.

Data Input

The application accesses raw data from any cloud storage service that is made available to it. The uniform resource locator (URL) of the data source is all that is required to be input. The application server is then commanded to retrieve the raw data and store it in the cloud-based application storage. Amazon S3 is utilised as the application's storage service. All data input and processed is backed-up and stored in this repository for later use in reporting and for replication purposes.

The data is retrieved and initial pre-processing completed by an *Rserve* endpoint.

The function firstly retrieves the data from the remote repository. It then prepares it for further processing by ensuring all data is of the required format. This data set is saved in the application storage repository (S3). Finally, the load profile of the dependent variable is plotted and returned to the user for validation purposes.

Feature Selection

Feature selection is a key and novel offering of IntelliMaV. It is required to maximise both model fit and the computational efficiency of applying machine learning algorithms. An adapted version of a novel feature selection algorithm presented in Chapter 4 is implemented in this M&V 2.0 application. This simple approach ranks input variables based on their importance and iteratively adds them to the feature set. An OLS regression model is trained at each stage and the $R^2_{adjusted}$, is used to compare model performance. A variable must increase the $R^2_{adjusted}$ by more than 5% to be considered a model feature and thus, added to the feature set. A 5% metric improvement threshold is employed to reduce the number of features selected. This is in contrast to the 1% threshold utilised in earlier research. There is scope to adjust this threshold in an iterative manner in future work. Algorithm 3 details the algorithm implemented in IntelliMaV. It is also important to note that the VIF is used to test for multi-collinearity between features. In keeping with the M&V principle of conservatism, any feature found to have a VIF greater than 5 is removed from the feature set.

Figure 6.2 shows an example of the user interface at the feature selection stage. An overview of the process is outlined to the user, who then must click to apply the algorithm. The task is executed on *Rserve*. The full feature set and relevant statistical measures are returned to the user upon successful completion of the task.

Data Cleaning

Data cleaning must be completed to ensure the data passed to the modelling algorithms are optimised for performance. Features with outliers, periods of missing data or unreliable measurements must be omitted from the feature set. As per the IPMVP, backfilling of data is not allowed, hence, the quantity of training data is often reduced as a result of the cleaning process.

On a high-level, any feature that is missing more than 5% of data points is omitted. For individual data points, an outlier detection method is applied to identify points that are numerically distant from the rest of the data. In typical box and

Algorithm 3: Wrapper-based feature selection that utilises the Spearman rank correlation and the adjusted coefficient of determination to optimise adjusted coefficient of determination

Input: m x n matrix containing all input data.
$x[\ ,n]$ = dependent variable
Apply Spearman rank correlation algorithm to calculate variable *ranks*
Order columns in input matrix by decreasing ρ
$i = 1$
ρ_i = Spearman correlation coefficient between variable i and $x[,n]$
$subset_i = x[\ ,cols(1,\ldots,i,n)]$, i.e. variable with highest ρ and dependent variable
Train OLS regression model for $subset_i$ & find $r^2_{adj_i}$
while i *!= no. of variables* **do**
    $subset_{i+1} = x[\ ,cols(1,\ldots,i+1,n)]$
    Train OLS regression model for $subset_{i+1}$ & find $r^2_{adj_{i+1}}$
    if $r^2_{adj_{i+1}} - r^2_{adj_i} > 0.05$ **then**
        $subset_i = subset_{i+1}$
        $r^2_{adj_i} = r^2_{adj_{i+1}}$
        $i = i + 1$
    end
    else
        Remove variable $i + 1$ from the data set
    end
    return $subset_i$
end
Output: data set with features selected.

whisker plots, an outlier is considered any value that is more than 1.5 times the inter quartile range (IQR) above the third quartile or below the first quartile. In IntelliMaV, a threshold of 3 times the IQR is implemented to account for the variability of energy data across the broad spectrum of operations. Outliers are removed from the data with no backfilling performed. Additionally, any feature found to consist of more than 10% outliers is removed from the feature set as removal of this data would reduce the training data set in size by too significant a factor. Section 6.4.1 includes an example of the implementation of the data cleaning process.

Development of Baseline Energy Model

The development of the baseline energy model has been identified as a critical step in M&V. Chapter 3 highlights the potential performance improvements possible when advanced machine learning techniques are applied. This is in direct comparison to the simplistic modelling techniques traditionally employed in

M&V. However, the computing resources required can act as a hindrance to the adoption of advanced modelling techniques. IntelliMaV utilises cloud computing infrastructure to overcome this problem.

Using the M&V 2.0 framework presented earlier, the model training process in IntelliMaV is implemented in line with the guidance documentation. Figure 6.3 illustrates the process followed to train the baseline energy models. An exhaustive process is used to ensure the model developed is tailored to the characteristics of each individual project.

Firstly, the clean feature set is aggregated for four different measurement frequencies; quarter-hourly (usually the base frequency), hourly, daily and weekly. This results in the creation of four data sets that facilitate an exhaustive approach to modelling. Z-score normalisation is then performed on each individual data set resulting in each feature having the properties of a standard normal distribution (i.e standard deviation of 1 and mean of 0). The data sets are then randomly partitioned into training and testing data sets using an 80:20 split ratio. The training data set for each measurement frequency is input into four different machine learning regression algorithms. These are OLS, k-NN, ANN and SVM. A grid-search approach and 10-fold cross validation are utilised to find the optimal values of each hyper-parameter. The hyper-parameters grid search values are those detailed in Chapter 4.

The process results in 16 models being trained with each evaluated on an unseen testing data set. It is important to note that this partitioning of data and testing on a previously unseen data set is not required by the IPMVP and ASHRAE guidance documentation. These approaches use 100% of the pre-ECM data to train a model and apply the model to the same data set for prediction performance evaluation. As stated previously, this approach is prone to over-fitting in cases with multiple degrees of freedom in the model. This can lead to presenting biased performance metrics, thus decreasing its usefulness outside of the baseline period.

The model training algorithm is implemented using the software architecture detailed in Figure 6.4. The user submits a HTTP request via the IntelliMaV front-end GUI. Following this, an API call is made on the web server to the *Pyserve*, which in turn directly communicates with AWS. Pyserve compiles the model training job which includes specifying the location of training data, allocating computing resources and specifying storage locations for process outputs. The training job is passed to Amazon Sagemaker, a server-less computing platform for machine learning, via an API call. The training code image is represented as

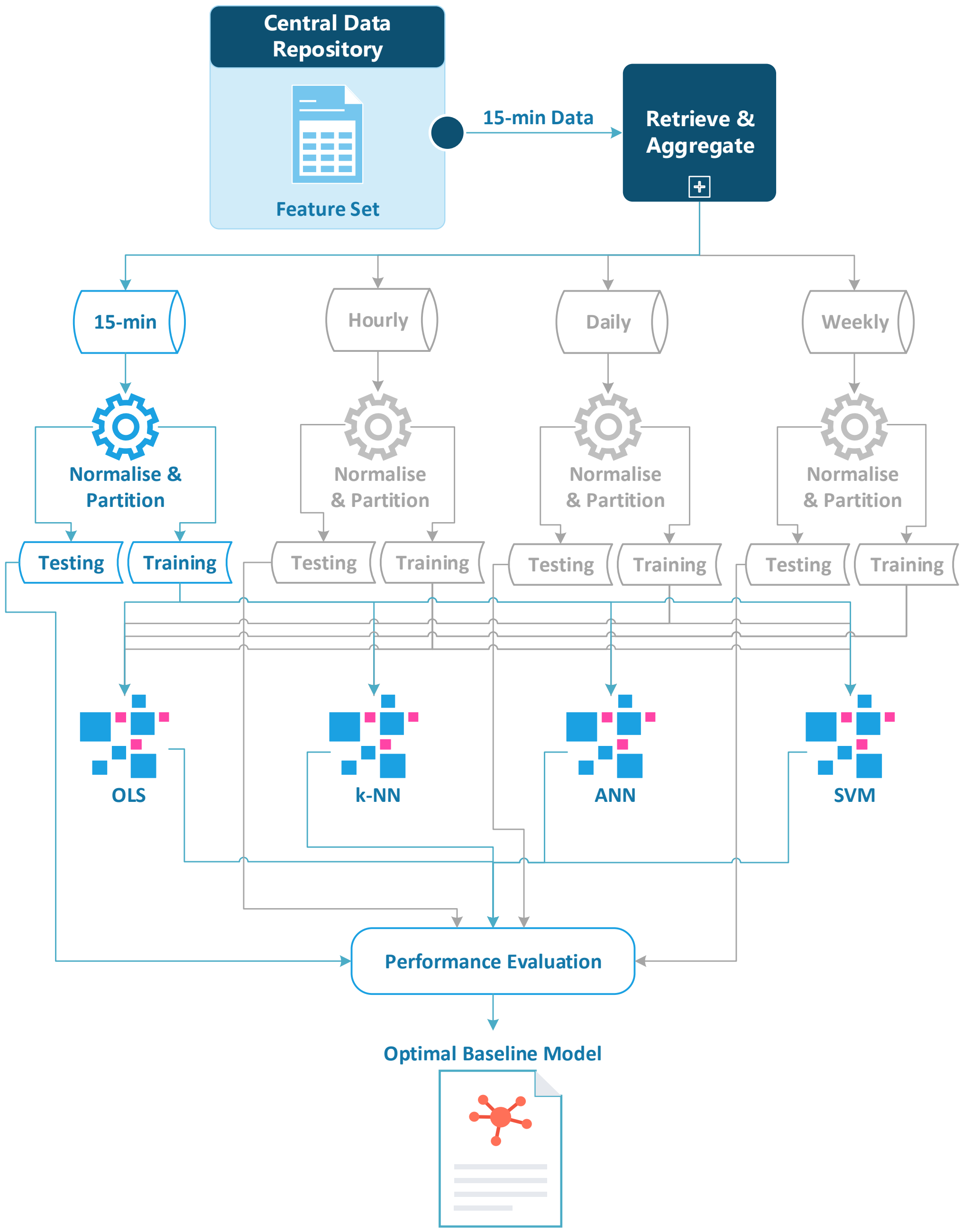


Figure 6.3: Flow diagram of model training process implemented in IntelliMaV

a Docker container image stored on the Amazon ECR. Docker is an open-source program that performs operating system level virtualisation. A Docker image is a file used to execute code in a Docker container. Containers are created from images which specify their precise contents. These containers are isolated and bundle applications in a lightweight environment. The image used in IntelliMaV contains code written in the R programming language for algorithm application.

The use of cloud computing infrastructure to train the baseline energy models

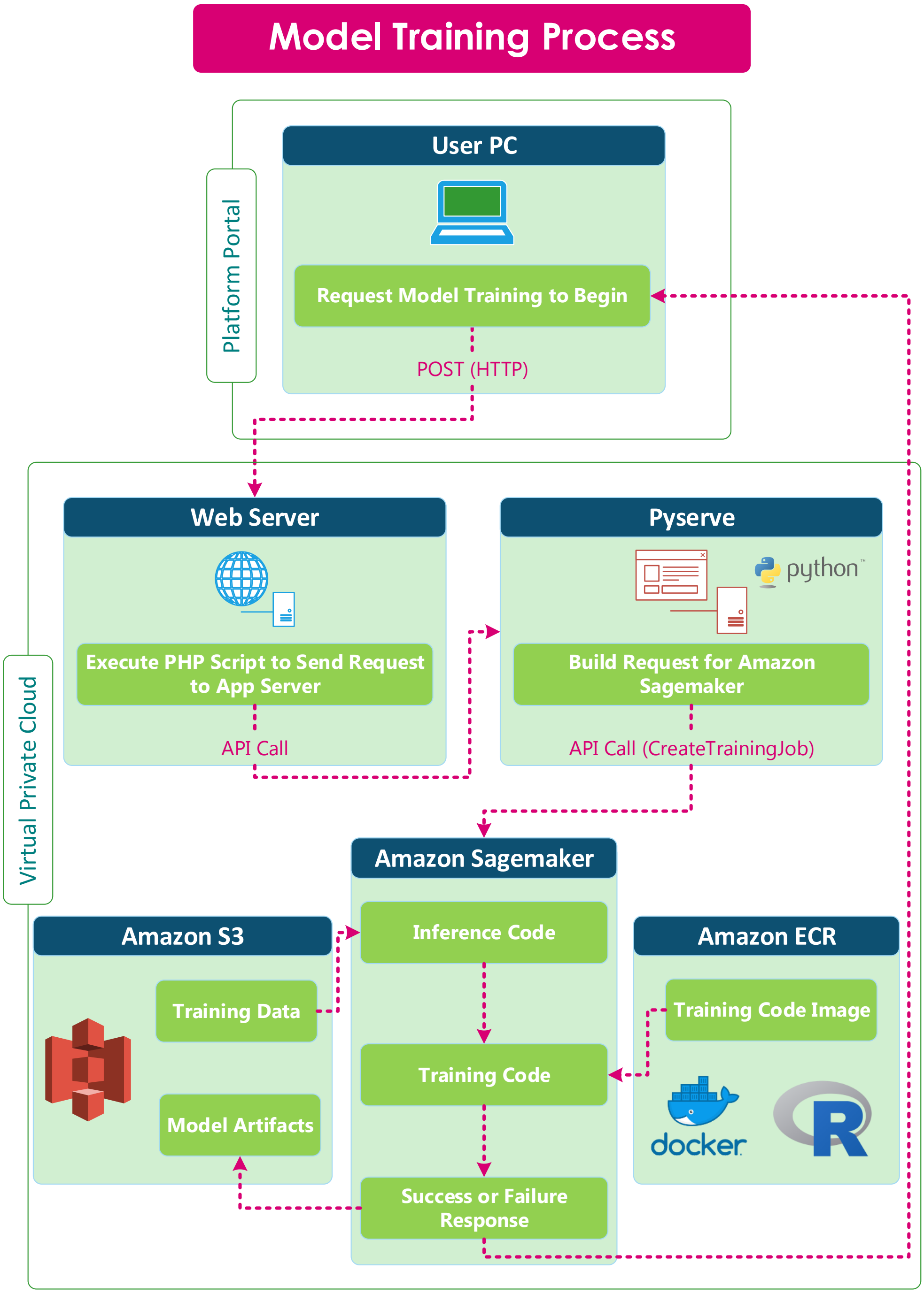


Figure 6.4: Overview of model training software architecture of IntelliMaV

offers the significant benefit of enabling all models to be trained in parallel using powerful computing resources. This reduces the model training time when compared to execution on a single machine. The processing time required to train machine learning algorithms on a data set can be a significant barrier to their implementation in M&V. This issue is exasperated when tasks are executed on a single personal computer. IntelliMaV efficiently executes these tasks on behalf of the practitioner, ensuring they no longer require powerful computing resources on-site. It allows the user to be thorough in their analysis, while maintaining efficiency and minimising costs.

6.3.3 Model Deployment

Model Evaluation

The model training process discussed results in the development of 16 baseline energy models. This number may be smaller depending on individual project characteristics as some algorithms may not fit to the data. Nevertheless, the optimal model must be identified using an evaluation procedure.

As has been discussed at length, the primary objective of M&V is to quantify energy savings with minimal uncertainty. Thus, this must be considered when developing a model evaluation procedure. Using the IPMVP guidance documentation defined approach to the calculation of uncertainty in M&V, the standard error, or RMSE, of the baseline energy model is found using Equation 6.1 (discussed in earlier chapters), where y_i is the measured value, $\hat{y}_i$ is the predicted value, n is the total number of data points and k is the number of independent variables or model features (Efficiency Valuation Organization 2018).

$$SE_{\hat{y}} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k-1}} \tag{6.1}$$

The coefficient of variation of RMSE, detailed in Equation 6.2 where $\bar{y}$ is the average of the measured values, is used to normalise the error metric with respect to the value of the measured quantity. This enables easier interpretation of results.

$$CV(RMSE) = \frac{SE_{\hat{y}}}{\bar{y}} * 100 \tag{6.2}$$

As the RMSE is directly used to calculate the uncertainty in the baseline energy model, it must be minimised to maximise accuracy in results. Thus, the optimal baseline energy model is evaluated based on this metric. As discussed in Section 6.3.2, the feature set is randomly split into training and testing data sets with the training data set being used to develop the baseline energy models. The models developed are then applied to the respective testing data sets and thus, the performance metrics are calculated. The use of a random data split and a testing data set aims to avoid models which have over-fit the data being identified as optimal. The 10-fold cross validation utilised in the model training algorithm also aids the reduction of this risk. The RMSE of each model fitted to the data set is calculated with the model producing the lowest error deemed to be the optimal baseline energy model. All other models are discarded at this point.

Deployment Parameters

The reporting period parameters must be defined prior to final deployment of the baseline energy model. The user must input the following details to enable the automated application of the model throughout the project life-cycle:

- URL of post-ECM data
- Reporting period start date
- Expected/estimated energy savings in kilowatts (kW)
- Description of implemented ECM and works that took place
- Photographs showing evidence of implementation for inclusion in final report (optional)

The maximum frequency with which the optimal baseline energy model can be applied is the measurement frequency of the data used to fit the algorithm. For example, an ANN algorithm fitted on daily consumption data can be applied daily at most.

Optimal Model Deployment

Following the definition of the post-ECM deployment parameters, the user begins the automated, near real-time energy savings quantification. Beginning on the user defined start date, a scheduled batch script runs on the web-server. This script makes an API call to the application server, '*Rserve*', which subsequently gathers data from the storage URL, applies the same cleaning as was applied pre-ECM and calculates energy savings and uncertainty using the optimal baseline energy model. As is common across all projects, the savings are calculated as being the difference between the measured energy consumption and the adjusted baseline estimated using the baseline energy model. Uncertainty is calculated at a 90% confidence interval. The results of the M&V process are visible to the user via a web dashboard.

6.3.4 Performance Deviation Detection System

The persistence period occurs outside the scope of traditional M&V. This is the point at which M&T takes over the evaluation of system performance and the process is an ongoing task for the duration of the ECMs lifetime. This is in contrast to typical M&V in which savings are quantified over a cycle of operation, usually 12 months, and extrapolated for the lifetime of the ECM. This new period of analysis enables performance evaluation to be an ongoing energy management task.

Prior to the implementation of an ECM, a Basis of Design (BOD) is often completed to assess the potential savings and associated costs. This will result in an estimation of performance. If this has not taken place, an engineering first-principles approach should be used to estimate the savings that will be achieved. Blomqvist et al. published a database of ECMs that can be used for estimating energy savings also (Blomqvist & Thollander 2015). This is a useful resource if no BOD has taken place.

In any case, the actual performance found using the baseline energy model is compared with the expected performance to establish if the savings are on track. Any deviations from expected performance triggers an exception report to the engineering team. As a rule of thumb, a 20% deviation is defined as a deviation from expected performance. This threshold was arrived at after considering the potential error in the preliminary estimation of savings used to compute it. Practitioners may chose to employ a lower threshold for stricter control. This automated system provides an insight into system performance, enabling corrective action to be taken to maximise the savings realised.

It is critical that the validity of the model is assessed on an ongoing basis. In line with criteria outlined in ASHRAE Guideline 14, the model can only be applied while the independent variables are no more than 110% of the maximum and no less than 90% of the minimum values of the same variables used for constructing the baseline energy model. If independent variables stray outside of these bounds, then the error metrics associated with them are no longer valid. The user is then notified of this and must return to the baseline energy model training process to initiate the construction of a more suitable model.

6.4 Case Study: Results and Discussion

The cloud computing-based IntelliMaV was applied to quantify the savings from an ECM carried out in a large biomedical manufacturing facility in Limerick, Ireland. The facility operates a continuous process on a 24/7 basis with an annual shut-down coinciding with Christmas holidays. The ECM was implemented on the AHUs in operation across the entire building. In total, AHUs account for approximately 8.5-9.5% of site electricity consumption. Prior to the implementation of the ECM, each AHU operated at a fixed speed with minimal control. The ECM consisted of the installation of instrumentation and software capable of implementing a proprietary control algorithm. The algorithm intelligently utilises VSDs, which were previously operated at fixed speeds, to vary the volume of treated air supplied to the building to meet the required environmental conditions. To be clear, IntelliMaV does not implement this optimisation algorithm as it is a performance verification tool. This is the same ECM that was analysed in Chapter 5, however, there are two keys differences. The baseline period data was restricted 12-months of data so as to ensure only data that represented the state of the energy system pre-ECM was included in analysis. In addition, the scope of analysis was expanded with an additional eight months of data available post-ECM. The same approach to verifying savings was taken with savings being assessed in the total electricity consumptions of all AHUs combined. It was also desired to utilise the existing M&T metering infrastructure on the site, thus avoiding the need to install additional meters and keeping project costs at a minimum. Figure 6.5 illustrates the energy consumption of the AHUs prior to the implementation of the ECM, i.e. the baseline energy consumption to be modelled.

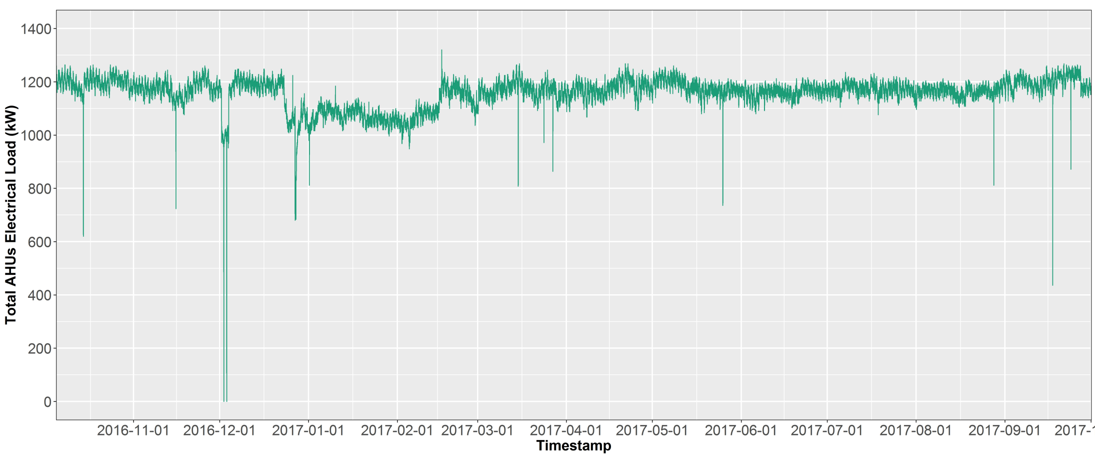


Figure 6.5: Electrical load of all AHUs in baseline period (pre-ECM)

6.4.1 Pre-ECM

Prior to the implementation of the ECM, IntelliMaV was used to develop a suitable baseline energy model. An initial project description and parameters were entered by the user. An assessment of available data revealed two primary data sources: the BMS and the EMS. All electrical, mechanical and climatic meters in place at the facility store data on a local server. This same server runs the BMS. This data is then made available using a web-based EMS. It was decided most appropriate to put in place a data pipeline between the local server and the secure, cloud storage used by the M&V 2.0 application. This automated the process of making local data available to the cloud-based IntelliMaV in a secure manner.

The local server contains historical data dating back to the beginning of 2016 for all meters on-site. All data have a base measurement frequency of 15-minutes. It is common practice in M&V to ensure a full cycle of operation is included in the baseline period. This is typically 12-months in duration for energy systems heavily influenced by climatic conditions. In production processes, this period generally consists of all operating stages. Industrial facilities present a challenge in that many variables affect the consumption of the energy system. In these cases, it is best practice to have a baseline period spanning 12-months, hence, the training data set was gathered using this approach. The baseline period started on 5$^{\text{th}}$ October 2016 and concluded on the 1$^{\text{st}}$ October 2017. It should be noted that this is in contrast to that 22 months of data employed to construct the baseline energy model in Chapter 5. Upon review of the findings of that case study, it was decided to limit the baseline period data to the most relevant full cycle of operation available. This was concluded after analysing the load profile of the energy system using the available data (Figure 6.6), in which it can be seen that in the early months of 2016 the energy systems behaves quite differently to the final 12 months. Despite the difference being small in the dependent variable, the changes in the independent variables in the same periods were large. Given the nature of the dynamic industrial facility and energy system, it was decided to discard these earlier data for the benefit of accuracy.

The data set made available contained 504 variables that are independent of the dependent variable, the AHUs electricity consumption. The feature selection algorithm discussed in Section 6.3.2 was applied to identify the most suitable independent variables to form the feature set. This task was completed efficiently by the application server which identified 7 highly correlated independent features.

Table 6.1: Results of feature selection process as returned to user

Feature	Description	Unit	Mean	No. of Unique	No. of Missing	Min	1^{st} Quartile	Median	3^{rd} Quartile	Max
MLP 1 Serving part of Phase 1	Electricity consumption in phase 1 of facility	kWh/15-min	46	36	197	0	44	46	49	67
HPN 2 Feeder	Feeder line	kWh/15-min	92	13	199	0	93	93	94	96
MDB E Incomer	Main electrical incomer E	kWh/15-min	59	207	207	0	22	23	24	323
PP30 37 IPA DB	Power panel 30	kWh/15-min	15	25	207	0	17	18	20	53
FI N50 01 Nitrogen Line 50	Flow in nitrogen line 50	l/s	0.6	12	234	0	0	0	0	11
MCC 50 WP	Motor control centre for WP area	kWh/15-min	40	16	220	0	39	39	41	45
Supply to Chiller 9	Electrical supply to chiller 9	kWh/15-min	13	105	454	0	0	9	20	472
Dependent Variable	The quantity directly impacted on by the ECM, i.e. the electricity consumption of all AHUs	kWh/15-min	285	144	0	0	282	292	298	330

Table 6.1 provides a description for each feature identified by the algorithm as being statistically relevant, while also including the summary statistics returned to the user.

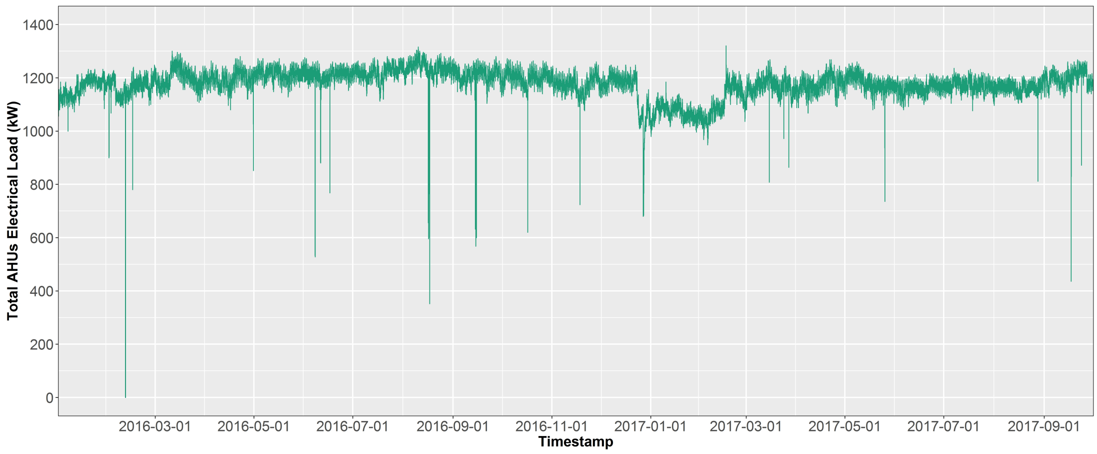


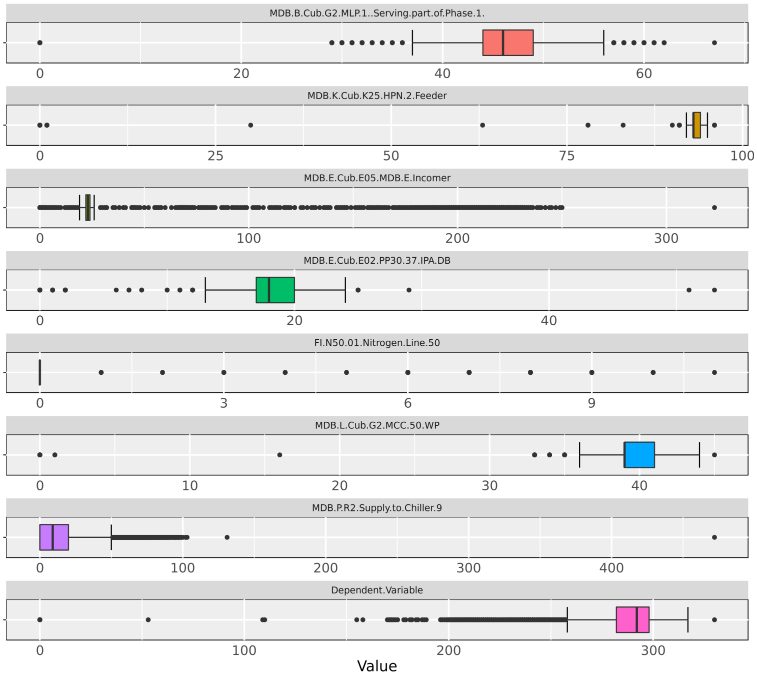
Figure 6.6: Electrical load of all AHUs in for duration of available data period

It is important to ensure data input to the regression algorithms are sufficiently clean as spurious data points can skew the model fit, thus reducing its validity post-ECM. Hence, the data cleaning process outlined in Section 6.3.2 was completed. This resulted in the feature set being reduced from 7 independent variables to 5. Figure 6.7 illustrates the omission of the two features and highlights the differences present in the data before and after cleaning. This process assists in achieving maximum accuracy in the model training process with the available data. Therefore, it ensures minimal error is introduced into the final savings quantified by the baseline energy model.

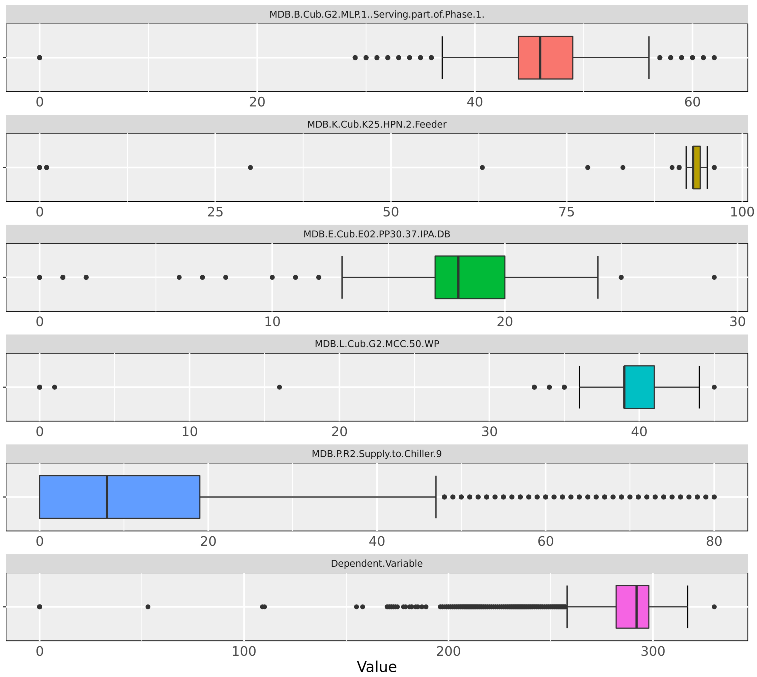
The cleaned feature set was then input into the model training process. This consists of the application of the four regression algorithms (OLS, k-NN, ANN and SVM) for four different measurement frequencies (quarter-hourly, hourly, daily and weekly.) The general equation of the desired model can be expressed in the form of the multiple OLS equation as follows:

$$AHU_{elec} = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5, \tag{6.3}$$

where b_0 is the value of AHU_{elec} when X_1 through to X_5 are equal to zero, B_{1-5} are the estimated regression coefficients and X_{1-5} are the variables in the final feature set. The application of the regression algorithms resulted in the construction of 15 baseline energy models. The k-NN algorithm failed to fit on the 15-minute data set.



(a) Feature set prior to data cleaning



(b) Feature set following data cleaning

Figure 6.7: Illustration of data cleaning process applied in the baseline period

6.4.2 Post-ECM

The model evaluation procedure discussed in Section 6.3.3 was used to identify the optimal baseline energy model to be deployed for post-ECM period. Figures 6.8 and 6.9 present graphical representations of the impact of temporal granularity and regression algorithm on the prediction performance of the models developed. In line with the overall objective of minimising uncertainty in the savings quantified, the model with the lowest RMSE was selected. For this case study, this was a k-NN model trained using data with an hourly measurement frequency.

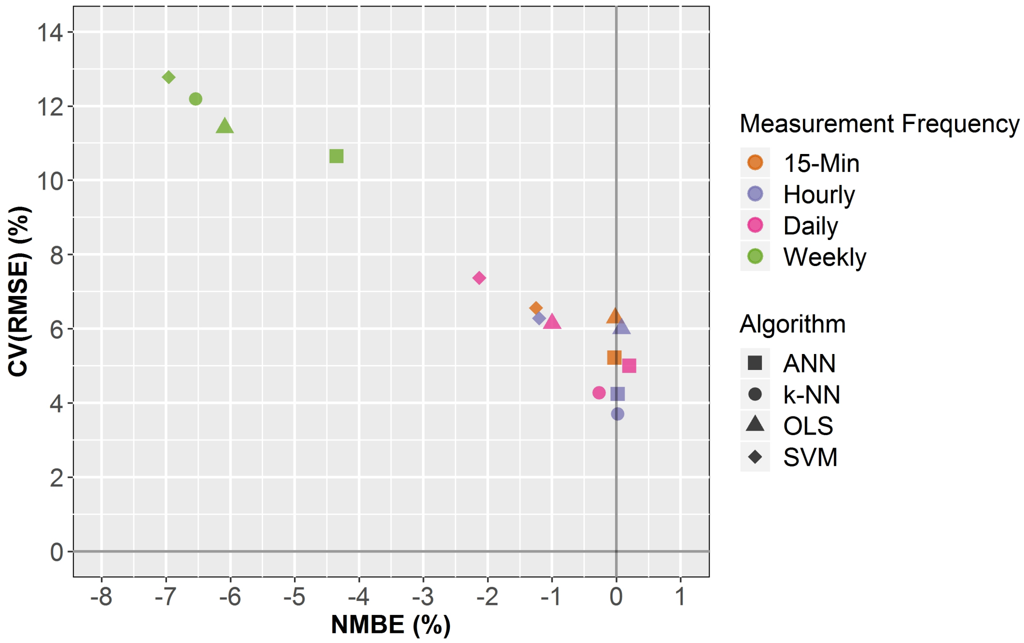


Figure 6.8: CV(RMSE) and NMBE for each model trained in baseline period

The reporting period parameters were then defined using the browser-based user interface. The reporting period began on 1st December 2017 and concluded on 9th August 2018. All installation and commissioning works had been fully completed by the beginning of the reporting period. A BOD that was carried out prior to the ECM's implementation estimated the savings to be 2.3 GWh per annum. This figure was input via the IntelliMaV user interface and subsequently used for performance deviation detection.

Following deployment of the optimal baseline energy model, IntelliMaV's dashboard was used to track savings in near real-time. Savings were estimated every hour with new measured data automatically uploaded to the cloud-based storage using the data pipeline implemented in the baseline period.

Section 6.3.4 details the PDD functionality of the application. A key step in this process is ensuring the model remains statistically valid for the duration of the ECM's lifetime. Figure 6.10 shows the model features both before and after implementation. In this case, the features remained within the acceptable bounds for the duration of the analysis. If the situation arose in which the features were

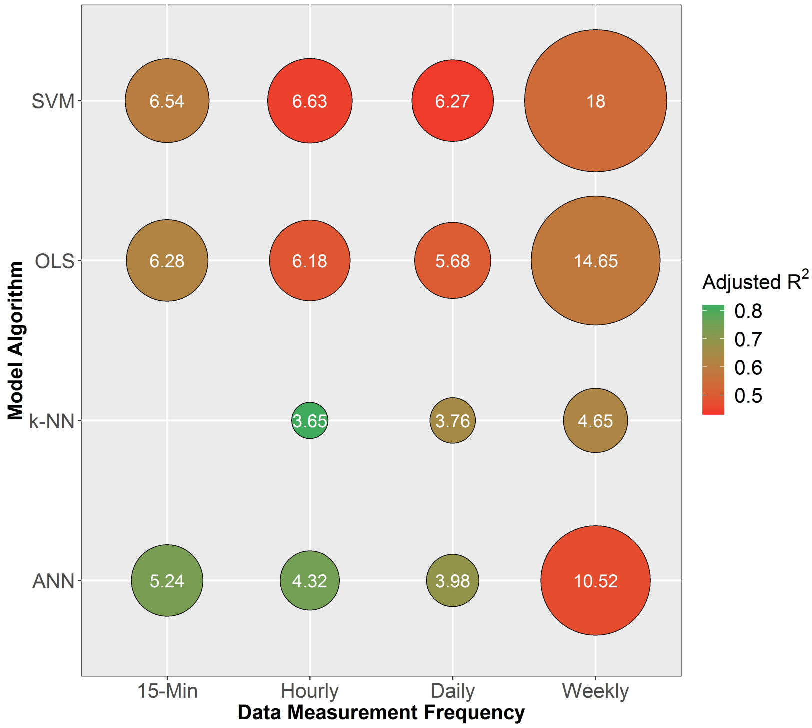


Figure 6.9: CV(RMSE) (bubble size and text) and associated r^2_{adj} (colour) for each model trained in the baseline period.

no longer valid with respect to their training values, then the model development process would need to be repeated.

The final energy savings were found to be 2,353,225 kWh/yr (268.6 kW) with 25.5% uncertainty at a 90% confidence interval. Therefore, it can be said with 90% confidence that the true value of the annualised savings lies within 1,749,372 kWh and 2,952,996 kWh. As per the IPMVP criteria, savings need to be larger than twice the standard error of the baseline model. The standard error, or RMSE, of the optimal model was found to be 10.4 kWh/15-min (41.6 kW). Therefore, the final energy savings were quantified with uncertainty at 31% of the maximum acceptable deemed level. This highlights the accuracy of the approach, thus increasing confidence in results.

6.5 Conclusions

This chapter presents a novel, cloud computing-based application that applies advanced machine learning techniques on large data sets to verify the performance of ECMs automatically and in near real-time. The primary barriers to the widespread adoption of machine learning techniques in the field of M&V are the

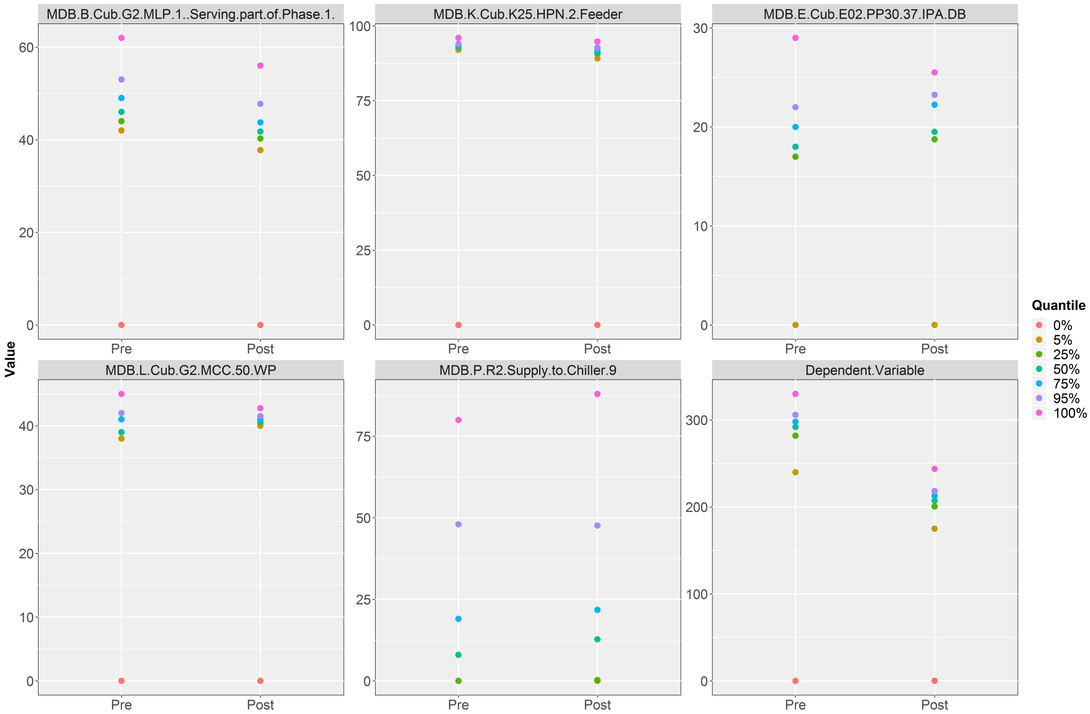


Figure 6.10: Comparison of feature percentiles before and after ECM implementation

computing resources and subject matter knowledge required to successful apply them. The utilisation of a cloud computing based infrastructure minimises the resources required on-site, while the robust functionality of IntelliMaV enables the user to apply the machine learning algorithms.

A PDD system is also presented to offer additional functionality to the application. The PDD system assesses actual realised energy savings with respect to the expected savings from an ECM. An exception report is issued to the user for any instances in which actual performance is more than 20% less than expected. This enables the M&V process to be more dynamic in nature and allows the practitioner to ensure savings persist over the lifetime of an ECM.

The success of commonly accepted M&V protocols, such as the IPMVP and ASHRAE Guideline 14, lies in their openness and applicability across the broad spectrum of ECMs. IntelliMaV was developed with this objective to the fore. This has been achieved through the use of a structured approach to the development of the baseline energy model, which requires the user to input project specific parameters. The baseline energy model is developed in a transparent, exhaustive manner with the optimal model being identified as that which minimises the RMSE. IntelliMaV is most useful when applied in the industrial buildings sector as these complex buildings generally require more in-depth analysis than the residential and commercial sectors. The automated nature of the application developed enables the task of M&V to be seamlessly integrated into ongoing energy management tasks. This allows M&V to progress to becoming an M&T task, thus ensuring energy savings persist long-term.

The M&V 2.0 application is capable of near real-time savings quantification. This demonstrates an evolution from the traditional static and retrospective analysis to the modern, dynamic and reactive approach that ensures persistence of energy efficiency savings. There are two limitations to the delivery of near real-time performance verification; the measurement frequency of the training data and the site specific data pipeline. The savings can only be quantified at a maximum of the same frequency as that of the measurement frequency of the data with which the optimal baseline model was trained. Additionally, IntelliMaV requires data to be made available to its cloud infrastructure. It is not practical to design a single solution to complete this task as there are many different data collection and storage configurations on industrial sites. This issue is mainly caused by legacy and proprietary systems in place. However, there is an opportunity to develop a solution based on emerging technologies and standards, such as the OPC Unified Architecture.

A case study was used to demonstrate the ease of use and effectiveness of the application developed. Real-world data from a large biomedical manufacturing facility in Ireland was used to measure and verify the savings resulting from the implementation of an intelligent control algorithm on the building's AHUs. The savings were quantified across all AHUs using the electricity consumption meters installed on each. The step-wise approach of IntelliMaV was then followed to successfully train a k-NN model on data with an hourly measurement frequency. The accuracy of this optimal model was evaluated on a previously unseen testing data set also gathered in the baseline period. The model was found to have a CV(RMSE) of 3.65%. The application was then capable of deploying this model in an automated fashion for the duration of the reporting period. The annualised savings were quantified as 2,353,225 kWh/yr with 25.5% uncertainty at a 90% confidence interval. M&V was successfully carried out with uncertainty 69% lower than the level deemed acceptable by the IPMVP.

Chapter 7

Conclusions

7.1 Summary of Research

This thesis focuses on advancing the practices used to measure and verify energy savings in industrial facilities. The significance of the uncertainty that is associated with any quantification of energy savings and it's potential impacts on the effectiveness of energy policy were highlighted in Chapter 2. Given that there are three sources of uncertainty in M&V (measurement, sampling and modelling), it was decided to focus attention on the uncertainty introduced in the construction of the baseline energy model. Thus, the techniques available to minimise the modelling uncertainty introduced to an M&V project are the focal point of the thesis. The traditional performance verification practices that rely on simplistic approaches often lack the power to successfully represent the state of complex energy systems. More intricate models must be developed to progress the performance verification sub-sector of energy management. In addition to improving the energy modelling techniques, the requirement for M&V 2.0 guidance documentation and solutions tailored for the industrial buildings sector were identified as opportunities to be pursued. Therefore, the research conducted can largely be classified into two fundamental tasks: data-driven energy modelling and a comprehensive M&V 2.0 solution.

The research detailed throughout this thesis begins with an assessment of the potential of machine learning techniques to more accurately model the complex energy systems that are in operation in modern industrial facilities. In a comparative analysis between an assumed typical M&V approach and that which employs more powerful, machine learning knowledge discovery algorithms, the latter were found to significantly improve the prediction accuracy of the baseline energy model. This directly results in the uncertainty in final energy savings being reduced. Additionally, it is important to note that the application of these regression algorithms on granular data sets reduces the spread of error across the models. This is significant in improving the reliability in a black-box approach. These findings offered promise, however, in isolation they do not aid the standard M&V practitioner in the adoption of such practices in real-world applications. To fill this void in knowledge, a machine learning-based energy modelling methodology was developed to provide prescriptive guidance on the use of these analytical techniques.

With the previous two chapters focusing on the energy modelling task in M&V, Chapter 5 presents an all-encompassing M&V 2.0 framework that builds on earlier research. The framework outlines the steps required to not just develop the baseline energy model, but also to deploy it using automated methods that will enable near real-time quantification of energy savings. In addition, findings from the research area of AFDD were leveraged to develop a simple, yet effective PDD system that seeks to maximise energy savings over the entire lifetime of an ECM. The body of work presented culminates in the development of a cloud computing application for performing M&V 2.0 in industrial buildings. IntelliMaV utilises the insightful findings of previous chapters to empower performance verification practitioners in the use of advanced data analytics techniques.

7.2 Success in Respect of Research Objectives

In Chapter 1 (Section 1.2), the universal objective of this research was defined as being the development of a data science-rooted, accurate, transparent and robust M&V 2.0 solution tailored for the industrial buildings sector. This was sought as a means to removing the barriers in place that impede investment in cost-effective energy efficiency improvement measures. These barriers include risk, hidden costs and uncertainty. Five ROs were subsequently developed to guide the work towards the successful realisation of the overarching objective. The achievement of these ROs is critically reviewed as follows:

RO1. Demonstrate the suitability of machine learning techniques to reduce the modelling uncertainty introduced into energy savings quantification with respect to current approaches.

The review of published literature in both the field of performance verification and the broader energy modelling research area identified black-box approaches as offering the most opportune manner with which to improve the techniques currently employed. This was concluded after a comprehensive review of white, grey and black-box approaches. Although grey-box approaches were found to offer promise in the future, they were deemed to lack the maturity of the data-driven black-box approaches that have been successfully employed in so many applications. To demonstrate the performance improvements that can be realised through the use of these data-driven approaches, a case study was conducted in a large, modern industrial facility. The case study facility was ideally suited to this type of assessment as no ECM had been performed there. Thus, it was possible to construct baseline energy models and review their prediction accuracy over a long period of time that is representative of a typical reporting period. This would not be possible had an ECM been implemented, as the reporting period (post-ECM) consumption would not allow for direct comparison with the baseline period.

The assumed typical M&V approach in the case study was that of a two-variable OLS regression model that employed outside air temperature and production electricity consumption as the predictors. This was deemed sensible based on industry knowledge and as both climatic and operational conditions are accounted for. However, it is commonly accepted that there are many different variables that impact on energy systems in these complex industrial facilities. Thus, the hypothesis was proposed that a broader scope of analysis would capture these relationships with greater accuracy than the assumed typical approach and subsequently, improve model performance. In the context of the case study, this was confirmed with the model CV(RMSE) and NMBE reduced by 15.9% and 75.6% respectively. This was achieved through the use of the OLS algorithm only. It was desired to quantify any further improvements possible through the use of the more advanced k-NN, ANN, decision tree and SVM approaches. A further 41.9% reduction in CV(RMSE) was achieved in this analysis. Finally, it was found using sensitivity analysis that the machine learning techniques are capable of overcoming one of the limitations of typical M&V in their ability to perform accurately with only limited training data available. It is important to note that Zhao & Magoulès (2012) concluded from a comprehensive review of the techniques available for energy modelling that no one technique can be identified as the best performer in every circumstance. A full review under a set of common conditions is required in each case. Thus, it is concluded that the case study sufficiently demonstrates the suitability of machine learning techniques to minimise the modelling uncertainty introduced into quantities of energy savings.

RO2. Investigate and develop data processing techniques tailored for the domain that ensure efficiency is maintained throughout the process, thus reducing the resources required to carry out M&V.

In Chapter 4, the development of a prescriptive energy modelling methodology that utilises machine learning techniques is presented. The pre-processing defined within this methodology includes algorithms that were developed with the industrial buildings sector in mind. The objective of effectively processing large data sets in a computationally efficient manner was used to guide the work undertaken. Thus, a novel wrapper method for feature selection is detailed. This makes use of the efficient Spearman rank-based correlation filter method, while also utilising OLS regression in the algorithm. This was deemed the most efficient manner to determine which variables in a data set contain sufficient knowledge to accurately represent the state of an energy system. A more complex solution would include each specific modelling algorithm (i.e. k-NN, ANN, SVM) in the feature selection process, however, this would be too computationally expensive due to the already resource intensive cross-validation process which cannot be omitted as it ensures robustness in the models developed.

As well as a problem specific feature selection algorithm, a data cleaning process is also presented. This provides clear guidance on the steps to take in ensuring noise is removed from the data set, while maintaining data integrity. These concepts were developed upon in Chapters 5 and 6, in which they were integrated into an M&V 2.0 framework and subsequently implemented in a final software solution. The ability of these data processing methods to ensure efficiency and effectiveness throughout were demonstrated in these chapters. These methods are required to ensure the modelling algorithms fit the data in the most accurate manner possible, while also minimising the resources required to carry out M&V through the focus placed on efficiency of processing.

RO3. Formalise a prescriptive methodology for the application of machine learning techniques to develop highly accurate baseline energy models for use in M&V 2.0.

Following on from the development of tailored and efficient data processing methods, these were used in the definition of a prescriptive energy modelling methodology that utilises machine learning techniques to construct highly accurate baseline energy models. The methodology was developed with a view to being implemented in M&V 2.0 applications, thus it was necessary that advanced analytical methods that can be automated were incorporated. The methodology presented is advantageous in circumstances with limited metering infrastructure, as powerful knowledge discovery algorithms are employed. In a case study, the energy savings resulting from a real-world ECM were quantified within acceptable uncertainty limits using the proposed approach. It is important to note that this case study presented challenging circumstances in which traditional M&V would not have been possible without the installation of additional metering infrastructure. Thus, this allowed for the robustness and accuracy of the methodology to be demonstrated, while also reducing the resources required to verify performance. Crucially, the knowledge gap on energy modelling that exists in the established M&V protocols is populated by the methodology detailed.

RO4. Integrate the energy modelling methodology into a comprehensive M&V 2.0 framework with a view to embedding performance verification deeper into best practice energy management to ensure persistence of savings over the lifetimes of ECMs.

Chapter 5 represents the beginning of the second task that this thesis focuses on. A comprehensive M&V 2.0 framework is developed to provide guidance to practitioners in the industrial buildings sector. This is of particular importance as the area is devoid of any guidance documentation, as proprietary offerings dominate the market. As with the IPMVP and ASHRAE Guideline 14, the principle of applicability across the broad spectrum of ECMs was at the forefront of the research. The framework presented integrates the baseline energy modelling methodology developed in Chapter 4, while remaining agnostic in terms of technology. This allows it to be applied in a number of different manners.

Critically, a new period of analysis is introduced to the performance verification process. The persistence period was formulated to enable the transition from traditional short-term M&V to long-term M&T as an ongoing task in good energy management. The use of a simple PDD system ensures savings persist over the duration of an ECM's lifetime. This rule-based approach represents the early stages implementation of such a system with the field of AFDD possessing many key findings that can be leveraged going forward.

RO5. Develop a computationally efficient and intelligent solution for near real-time energy savings quantification and performance deviation detection.

Chapter 6 represents the culmination of the research conducted and detailed in the previous chapters. A novel, cloud computing-based application that applies advanced machine learning techniques on large data sets to accurately verify the performance of ECMs in an automated fashion and in near real-time is presented. IntelliMaV implements the M&V 2.0 framework detailed in Chapter 5, which in turn incorporated findings from previous works. The automated nature of the application enables the traditionally short-term (i.e 1 to 2 years) tasks of performance verification to be seamlessly integrated with ongoing M&T. This allows for savings to be tracked for the long-term duration of an ECM's lifetime. IntelliMaV was applied to quantify the savings resulting from an ECM in a large, biomedical manufacturing facility. This results in savings being quantified with just 17.5% uncertainty at a 90% confidence interval. The case study application demonstrates the ease of use and accuracy of the solution developed. IntelliMaV is not only evidence of achieving RO5, but also the realisation of the universal RO defined at the beginning of this thesis.

7.3 Critical Appraisal of Research Undertaken

The scope of the research detailed in this thesis is defined from the beginning in Chapter 1 (Section 1.3). In this, it states that the findings are mostly limited to the field of performance verification. Despite this, there are conclusions drawn in Chapter 3 that have relevance beyond the scope of M&V to the broader field of energy modelling. As a whole, it is the opinion of the author that the work undertaken is effective in addressing the barriers to investment that M&V practices can be responsible for. Thus, the findings contribute to the closing of the energy efficiency gap. This is a critical step in ensuring the energy efficiency resource is utilised to its maximum potential as the transition towards low-carbon economies takes place. The findings not only contribute to the knowledge base in the research area, but also provide guidance that facilitate action to be taken in end-use energy systems. However, it is also important to highlight the limitations of the work undertaken. As the work undertaken can be classified into two high-level tasks, this classification is used in conducting a critical appraisal of the work.

7.3.1 Chapters 3 and 4: Modelling Uncertainty

The findings of a review of published literature dictated the initial direction of the analysis. Thus, the objective of assessing the suitability of black-box machine learning techniques to improve the performance of the baseline energy model in M&V was defined. Previous successes in the broader research area of data-driven energy modelling and the advent of M&V 2.0 requiring more advanced analytics were the instigators of the work detailed in Chapter 3. In this work, four modelling algorithms, k-NN, ANN, SVM and decision trees, were assessed along with the assumed typical OLS approach. It is important to note that there is scope for this analysis to be broadened through the use of other modelling techniques. In addition to applying different modelling algorithms, the grid-search parameters applied could also be evolved. The values utilised in the analysis were selected in an effort to maximise robustness in the approach.

In addition to this, there were further decisions made in Chapter 4 in the development of a methodology that is capable of utilising these powerful modelling algorithms. One such decision was the use of the VIF to identify multi-collinearity (or variable interdependence) and the subsequent definition of a threshold value for variable omission. Again, this value was arrived at following trial and error and after consulting published literature. There is however scope to improve on this process to ensure the most appropriate threshold is defined. There is also the opportunity to utilise PCA to tackle multicollinearity in the methodology.

The use of a random data split to quantify the error of each baseline energy model was discussed at length. It was highlighted on more than one occasion that the approach recommended by the IPMVP is prone to over-fitting. The random data split was used to overcome this problem and the results were found to be reliable. Despite this, a more advanced solution would utilise bootstrapping techniques in the calculation of uncertainty.

Finally, the case study presented in Chapter 4 contained challenging circumstances. The benefits of this have been publicised as aiding the methodology development process in improving the robustness of the final solution. However, the results highlight the limitations of the methodology in that it is dependent on the quality of data input. The range of savings in the case study were found to be 256,485-952,568 kWh at a 68% confidence interval. Although this is acceptable under the criteria defined by EVO and ASHRAE, it also suggests that the acceptability criteria need to be made more stringent. As discussed in Chapter 2, the impact of uncertainty in individual projects on the effectiveness of national energy policy is significant and thus, every effort must be made to improve the processes in place.

7.3.2 Chapters 5 and 6: M&V 2.0 Solution

Chapters 5 and 6 present work undertaken to develop an M&V 2.0 solution tailored to the industrial buildings sector. This need was identified due to the research field and markets being devoid of any guidance documentation or sufficiently powerful software offerings. In Chapter 5, an early stage case study was carried out based on the data available at that time. Thus, the reporting period in this case only spans 19.5 days. The same ECM was used in Chapter 6, however, a progressed version of the case study was presented as an additional 8 months of data was available post-ECM.

It was noted in the individual chapters that there are three key differences between the methods applied in these case studies. One is that of the longer reporting period, while the baseline period chosen for analysis also differs. In Chapter 5, a longer baseline period was employed with 22 months of training data gathered. This is useful in that the modelling algorithms generally perform better in data rich environments. Despite this, a judgement must be made in each individual application of the M&V 2.0 framework developed. That is, how much of the baseline data is representative of the energy system's state just before the ECM is implemented? Given that large industrial facilities can be ever changing environments due to the complexity of factors that impact on energy consumption, it was decided in Chapter 6 that only 12 months of baseline data be employed.

The final difference is the threshold used in the feature selection algorithm. In Chapter 5, a 1% improvement in $R^2_{adjusted}$ was used as the threshold for deciding on which variables were worth employing for model development. This resulted in 18 features being selected, with the optimal model having 15.99 kW of standard error. In Chapter 6, a 5% threshold was used in the IntelliMaV application. 7 variables were selected using this approach, while subsequent data cleaning reduced this number of features to 5. The standard error of the optimal in this more comprehensive case study was 41.6 kW.

The impact on results of the three critical differences between both versions of the case study raises some interesting topics for discussion. The rationale for selecting a shorter baseline period in Chapter 6 was based on the theory of using the most recent cycle of operations to represent the energy system. Given that this resulted in higher standard error in the baseline energy model, there is a trade-off between the accuracy of the model fit and the relevance of the data used to construct it. Additionally, the higher threshold used in the feature selection algorithm reduced the number of features selected from 18 to 7. The higher number of features in Chapter 5 resulted in a lower standard error in the baseline energy model. However, as each feature has an associated metering uncertainty, the uncertainty introduced by this model will generally be higher than that with the lower number of features. Hence, there is a balance that must be struck between the number of variables used to represent the energy system and the metering uncertainty introduced. Sensitivity analysis of the impacts of both the threshold used in feature selection and the length of the baseline period is required to fully understand this relationship. This however is not practical in each and every application. Thus, a standardised approach must be adopted in the industry.

The importance of a reporting period encompassing a full period of operation must also be stressed. In Chapter 5, with the 22-month baseline period and 19.5 day reporting period, the savings were quantified to be 380.1 +/- 26.4 kW. These were the savings realised throughout the reporting period only. The complex behaviour of the energy system under analysis throughout the year means that linearly extrapolating these savings to a full year of operation is not accurate. The results of Chapter 6 prove this hypothesis, with a 12-month baseline period and 8-month reporting period resulting in savings being quantified to be 268.6 +/- 68.5 kW. In both cases, a 90% confidence interval was employed. The variance in the energy consumption of the physical system and the times at which savings are realised throughout the year do not allow for linear extrapolation of savings. This is further evidence of the need for a performance verification system that continually verifies the performance of an ECM over its lifetime. This is in contrast to the retrospective and short-term approaches employed in the sector presently.

Above all, the impacts of the decisions made on a per-project level cannot be overlooked. The application of two varying approaches to the same case study in both chapters highlights these impacts. The baseline energy modelling methodology and M&V 2.0 framework are significant outputs from this thesis as they aid the standardisation of the methods employed. However, they are still dependent on the user to define a number of project parameters such as baseline and reporting period dates. A more comprehensive approach would assess the sensitivity of the final savings to changes in the values of these parameters. This would further enhance the reliability and trust in a practice that is not an exact science.

7.4 Recommendations for Future Research

This thesis presents a number of works which are intended to advance the practices in the field of performance verification of energy efficiency measures. The following research topics logically follow, based on the results presented in this body of work:

- There is an opportunity to assess the use of more complex machine learning techniques to further improve the accuracy of baseline energy modelling. Such approaches would include ensemble methods and deep learning. Deep learning is a subset of the field of machine learning and is receiving a lot of attention in recent times due to its ease of implementation made possible by advancements in computing power.
- The scope of this thesis is limited to the modelling uncertainty in performance verification. An assessment of the impact of measurement uncertainty using the proposed approaches would be beneficial. This is of particular relevance as more independent variables are being employed with respect to traditional M&V methods. Generally, each independent variable represents at least one physical meter, thus the resultant changes in measurement uncertainty would provide further insight into the process.
- A simple, rule-based PDD system is employed to identify degradation of system performance. This is effective in introducing this practice into the M&V process. Despite this, the mature research area of AFDD consists of many more advanced approaches that should be leveraged to improve the system in performance verification.
- In addition to advancing the PDD system, there is also an opportunity to further integrate IntelliMaV with widely used M&T tools. This is achievable by utilising the Predictive Model Markup Language (PMML) to describe and exchange the predictive models and results developed with other applications. Additionally, the utilisation of the *Haystack* naming convention facilitates interaction between data rich systems. Success in this quest will aid the movement of M&V away from a standalone task to a key tool in continuous energy management.
- As mentioned in Section 7.3, the reliability and trust in the performance verification process could be further improved by performing sensitivity analysis on the impacts of changes in baseline and reporting period lengths on final energy savings and uncertainty.

This thesis addresses some of the key challenges facing the performance verification industry at present. The implications of such challenges have been shown to be significant beyond the individual project level with the effectiveness of European energy policy reliant on accurate and reliable M&V. The methodology, framework and application developed all address these challenges, while aiding the transition to M&V 2.0 practices. Despite these advancements, this is not the final solution for the industry. A continued effort must be made to modernise performance verification practices to ensure M&V remains a valued practice in energy management into the future.

References

Ahmad, A. S., Hassan, M. Y., Abdullah, M. P., Rahman, H. A., Hussin, F., Abdullah, H. & Saidur, R. (2014), 'A review on applications of ANN and SVM for building electrical energy consumption forecasting', *Renewable and Sustainable Energy Reviews* **33**, 102–109.

Allcott, H. & Greenstone, M. (2012), 'Is there an 'energy gap'?', *Journal of Economics Perspectives* **26**(4), 25.

Amasyali, K. & El-Gohary, N. M. (2018), 'A review of data-driven building energy consumption prediction studies'.

American Council for an Energy Efficieny Economy (2018), 'Energy efficiency as a resource'.

American Society of Heating Refrigerating and Air-Conditioning Engineers (2014), 'Guideline 14 - Measurement of Energy, Demand, and Water Savings'.

Antunes, P., Carreira, P. & Mira da Silva, M. (2014), 'Towards an energy management maturity model', *Energy Policy* **73**, 803–814.

Arens, M., Neef, C., Beckert, B. & Hirzel, S. (2018), 'Perspectives for digitising energy-intensive industries - Findings from the European iron and steel industry', pp. 259–268.

Backlund, S., Thollander, P., Palm, J. & Ottosson, M. (2012), 'Extending the energy efficiency gap', *Energy Policy* **51**, 392–396.

Basak, D., Pal, S. & Chandra Patranabis, D. (2007), 'Support vector regression', *Neural Information Processing - Letters and Reviews* **11**(10), 203–224.

Blomqvist, E. & Thollander, P. (2015), 'An integrated dataset of energy efficiency measures published as linked open data', *Energy Efficiency* **8**(6), 1125–1147.

Bruton, K., Coakley, D., Raftery, P., Cusack, D. O., Keane, M. M. & O'Sullivan,

D. T. (2015), 'Comparative analysis of the AHU InFO fault detection and diagnostic expert tool for AHUs with APAR', *Energy Efficiency* **8**(2), 299–322.

Bruton, K., Raftery, P., Kennedy, B., Keane, M. M. & O'Sullivan, D. T. J. (2014), 'Review of automated fault detection and diagnostic tools in air handling units', *Energy Efficiency* **7**(2), 335–351.

Bunse, K., Vodicka, M., Schönsleben, P., Brülhart, M. & Ernst, F. O. (2011), 'Integrating energy efficiency performance in production management - Gap analysis between industrial needs and scientific literature', *Journal of Cleaner Production* **19**(6-7), 667–679.

Burman, E., Mumovic, D. & Kimpian, J. (2014), 'Towards measurement and verification of energy performance under the framework of the European directive for energy performance of buildings', *Energy* **77**, 153–163.

Cagno, E. & Trianni, A. (2013), 'Exploring drivers for energy efficiency within small- and medium-sized enterprises: First evidences from Italian manufacturing enterprises', *Applied Energy* **104**, 276–285.

Capehart, B. L., Turner, W. C. & Kennedy, W. J. (2011), *Guide to Energy Management*, seventh ed edn, The Fairmont Press, Lilburn.

Capozzoli, A., Piscitelli, M. S., Brandi, S., Grassi, D. & Chicco, G. (2018), 'Automated load pattern learning and anomaly detection for enhancing energy management in smart buildings', *Energy* **157**, 336–352.

Capros, P., De Vita, A., Tasios, N., Siskos, P., Kannavou, M., Petropoulos, A., Evangelopoulou, S., Zampara, M., Papadopoulos, D., Nakos et al., C., Paroussos, L., Fragiadakis, K., Tsani, S., Karkatsoulis et al., P., Fragkos, P., Kouvaritakis et al., N., Höglund-Isaksson, L., Winiwarter, W., Purohit, P., Gomes-Sanabria, A., Frank, S., Forsell, N., Gusti, M., Havlík, P., Obersteiner, M., Witzke, H. P. & Kesting, M. (2016), EU Reference Scenario 2016: Energy, Transport and GHG emissions trends to 2050, Technical report.

Carbon Trust (2018), 'Energy management self-assessment tool'.

Carpenter, J., Woodbury, K. A. & O'Neill, Z. (2018), 'Using change-point and Gaussian process models to create baseline energy models in industrial facilities: A comparison', *Applied Energy* **213**, 415–425.

Carpenter, J., Woodbury, K. & O'Neill, Z. (2016), 'A Comparison of Gaussian

Process Regression and Change-Point Regression for the Baseline Model in Industrial Facilities', *IBPSA-USA Journal* **6**(1).

Carstens, H. (2017), A Bayesian approach to energy monitoring optimization, PhD thesis.

Carstens, H., Xia, X. & Yadavalli, S. (2017*a*), 'Efficient metering and surveying sampling designs in longitudinal Measurement and Verification for lighting retrofit', *Energy and Buildings* **154**, 430–447.

Carstens, H., Xia, X. & Yadavalli, S. (2017*b*), 'Low-cost energy meter calibration method for measurement and verification', *Applied Energy* **188**, 563–575.

Carstens, H., Xia, X. & Yadavalli, S. (2018*a*), 'Bayesian energy measurement and verification analysis', *Energies* **11**(2), 380.

Carstens, H., Xia, X. & Yadavalli, S. (2018*b*), 'Measurement uncertainty in energy monitoring: Present state of the art', *Renewable and Sustainable Energy Reviews* **82**, 2791–2805.

Chiang, L., Russell, E. & Braatz, R. (2001), *Fault detection and diagnosis in industrial systems*.

Chou, J.-S. & Bui, D.-K. (2014), 'Modeling heating and cooling loads by artificial intelligence for energy-efficient building design', *Energy and Buildings* **82**, 437–446.

Coakley, D., Raftery, P. & Keane, M. (2014), 'A review of methods to match building energy simulation models to measured data'.

Council of the European Union (2007), 'Presidency Conclusions Brussels European Council 8/9 March 2007'.

de Wilde, P., Martinez-Ortiz, C., Pearson, D., Beynon, I., Beck, M. & Barlow, N. (2013), 'Building simulation approaches for the training of automated data analysis tools in building energy management', *Advanced Engineering Informatics* **27**(4), 457–465.

Díaz, J. A., Ramos, J. S., Delgado, M. C. G., García, D. H., Montoya, F. G. & Domínguez, S. Á. (2018), 'A daily baseline model based on transfer functions for the verification of energy saving. A case study of the administration room at the Palacio de la Madraza, Granada', *Applied Energy* **224**, 538–549.

Dong, B., Cao, C. & Lee, S. E. (2005), 'Applying support vector machines to predict building energy consumption in tropical region', *Energy and Buildings* **37**(5), 545–553.

Du, Z., Fan, B., Jin, X. & Chi, J. (2014), 'Fault detection and diagnosis for buildings and HVAC systems using combined neural networks and subtractive clustering analysis', *Building and Environment* **73**, 1–11.

Edwards, R. E., New, J. & Parker, L. E. (2012), 'Predicting future hourly residential electrical consumption: A machine learning case study', *Energy and Buildings* **49**, 591–603.

Efficiency Valuation Organization (2012), Measurement and Verification Protocol: Concepts and Options for Determining Energy and Water Savings Volume 1, Technical Report January, Efficiency Valuation Organization.

Efficiency Valuation Organization (2014), Statistics and Uncertainty for IPMVP, Technical Report June, Efficiency Valuation Organization.

Efficiency Valuation Organization (2016), Core Concepts, Technical Report October, Efficiency Valuation Organization.

Efficiency Valuation Organization (2018), 'Uncertainty Assessment for IPMVP'.

Efron, B. & Tibshirani, R. J. (1993), *An introduction to the bootstrap*, Chapman and Hall.

Ekici, B. B. & Aksoy, U. T. (2009), 'Prediction of building energy consumption by using artificial neural networks', *Advances in Engineering Software* **40**(5), 356–362.

European Commission (2018), 'Statement/18/4155'.

European Parliament and Council (2012), 'Directive 2012/27/EU of the European Parliament and of the Council of 25 October 2012 on energy efficiency'.

Eurostat (2016), 'Europe in figures - Eurostat yearbook 2016'.

Fan, C., Xiao, F. & Zhao, Y. (2017), 'A short-term building cooling load prediction method using deep learning algorithms', *Applied Energy* **195**, 222–233.

Finnerty, N., Sterling, R., Coakley, D. & Keane, M. M. (2018), 'An energy management maturity model for multi-site industrial organisations with a global presence', *Journal of Cleaner Production* **167**, 1232–1250.

Font Vivanco, D., Kemp, R. & van der Voet, E. (2016), 'How to deal with the rebound effect? A policy-oriented approach', *Energy Policy* **94**, 114–125.

Foucquier, A., Robert, S., Suard, F., Stéphan, L. & Jay, A. (2013), 'State of the art in building modelling and energy performances prediction: A review', *Renewable and Sustainable Energy Reviews* **23**, 272–288.

Franconi, E. M., Gee, M., Goldberg, M. L., Granderson, J., Guiterman, T., Li, M. & Smith, B. A. (2017), 'The Status and Promise of Advanced M&V: An Overview of "M&V 2.0" Methods, Tools, and Applications'.

Fumo, N., Mago, P. & Luck, R. (2010), 'Methodology to estimate building energy consumption using EnergyPlus Benchmark Models', *Energy and Buildings* **42**(12), 2331–2337.

Gallagher, C. V., Bruton, K., Leahy, K. & O'Sullivan, D. T. J. (2018), 'The suitability of machine learning to minimise uncertainty in the measurement and verification of energy savings', *Energy and Buildings* **158**, 647–655.

Gallagher, C. V., Bruton, K. & O'Sullivan, D. T. J. (2016), Utilising the cross industry standard process for data mining to reduce uncertainty in the measurement and verification of energy savings, *in* 'Lecture Notes in Computer Science Vol 9714', Springer International Publishing, pp. 48–58.

Gallagher, C. V., Leahy, K., O'Donovan, P., Bruton, K. & O'Sullivan, D. T. (2019), 'IntelliMaV: A cloud computing measurement and verification 2.0 application for automated, near real-time energy savings quantification and performance deviation detection', *Energy and Buildings* **185**, 26–38.

Gallagher, C. V., Leahy, K., O'Donovan, P., Bruton, K. & O'Sullivan, D. T. J. (2018), 'Development and application of a machine learning supported methodology for measurement and verification (M&V) 2.0', *Energy and Buildings* **167**, 8–22.

Gallagher, C. V., O'Donovan, P., Leahy, K., Bruton, K. & O'Sullivan, D. T. J. (2018), From M&V to M&T: An artificial intelligence-based framework for real-time performance verification of demand-side energy savings, *in* 'International Conference on Smart Energy Systems and Technologies', IEEE, Sevilla, Spain, pp. 1–6.

Ginestet, S. & Marchio, D. (2010), 'Retro and on-going commissioning tool applied to an existing building: Operability and results of IPMVP', *Energy* **35**(4), 1717–1723.

Golden, A., Woodbury, K., Carpenter, J. & O'Neill, Z. (2017), 'Change point and degree day baseline regression models in industrial facilities', *Energy and Buildings* **144**, 30–41.

Gopalakrishnan, B., Ramamoorthy, K., Crowe, E., Chaudhari, S. & Latif, H. (2014), 'A structured approach for facilitating the implementation of ISO 50001 standard in the manufacturing sector', *Sustainable Energy Technologies and Assessments* **7**, 154–165.

Granderson, J. & Fernandes, S. (2017), 'The state of advanced measurement and verification technology and industry application', *The Electricity Journal* **30**(8), 8–16.

Granderson, J. & Price, P. N. (2014), 'Development and application of a statistical methodology to evaluate the predictive accuracy of building energy baseline models', *Energy* **66**(1), 981–990.

Granderson, J., Touzani, S., Custodio, C., Sohn, M. D., Jump, D. & Fernandes, S. (2016), 'Accuracy of automated measurement and verification (M&V) techniques for energy savings in commercial buildings', *Applied Energy* **173**, 296–308.

Granderson, J., Touzani, S., Fernandes, S. & Taylor, C. (2017), 'Application of automated measurement and verification to utility energy efficiency program data', *Energy and Buildings* **142**, 191–199.

Greening, L. A., Greene, D. L. & Difiglio, C. (2000), 'Energy efficiency and consumption - the rebound effect - a survey', *Energy Policy* **28**(6-7), 389–401.

Guo, J. J., Wu, J. Y. & Wang, R. Z. (2011), 'A new approach to energy consumption prediction of domestic heat pump water heater based on grey system theory', *Energy and Buildings* **43**(6), 1273–1279.

Guyon, I. & Elisseeff, A. (2003), 'An Introduction to Variable and Feature Selection', *Journal of Machine Learning Research* **3**, 1157–1182.

Haberl, J. S. & Thamilseran, S. (1996), 'Great Energy Predictor Shootout II: Measuring retrofit savings - overview and discussion of results', *ASHRAE Transactions* **102**(2), 419–435.

Harish, V. & Kumar, A. (2016), 'A review on modeling and simulation of building energy systems', *Renewable and Sustainable Energy Reviews* **56**, 1272–1292.

Hastie, T., Tibshirani, R. & Friedman, J. (2009), *The Elements of Statistical Learning*, Vol. 18, Springer New York.

Henderson, H. & Velleman, P. (1981), 'Building Multiple Regression Models Interactively', *Biometrics* **37**(2), 391–411.

Heo, Y. & Zavala, V. M. (2012), 'Gaussian process modeling for measurement and verification of building energy savings', *Energy & Buildings* **53**, 7–18.

Hoffman, I., Schiller, S. R., Todd, A., Billingsley, M., Goldman, C. & Schwartz, L. (2015), Energy Savings Lifetimes and Persistence: Practices, Issues and Data, Technical report, Berkeley Lab.

Horowitz, M. J. (2011), 'Measuring the savings from energy efficiency policies: A step beyond program evaluation', *Energy Efficiency* **4**(1), 43–56.

Hu, L. Y., Huang, M. W., Ke, S. W. & Tsai, C. F. (2016), 'The distance function effect on k-nearest neighbor classification for medical datasets', *SpringerPlus* **5**(1), 1304.

Inard, C., Bouia, H. & Dalicieux, P. (1996), 'Prediction of air temperature distribution in buildings with a zonal model', *Energy and Buildings* **24**(2), 125–132.

International Energy Agency (2017), Energy Efficiency 2017, Technical report.

International Energy Agency (2018), World energy balances: Overview, Technical report, International Energy Agency.

International Organization for Standardization (2011), 'ISO 50001:2011'.

International Organization for Standardization (2014), 'ISO 50015:2014'.

International Organization for Standardization (2015), 'ISO 17742:2015'.

International Organization for Standardization (2016*a*), 'ISO 17741:2016'.

International Organization for Standardization (2016*b*), 'ISO 50047:2016'.

International Organization for Standardization (2018), 'ISO 50001:2018'.

Introna, V., Cesarotti, V., Benedetti, M., Biagiotti, S. & Rotunno, R. (2014), 'Energy Management Maturity Model: An organizational tool to foster the continuous reduction of energy consumption in companies', *Journal of Cleaner Production* **83**, 108–117.

Isermann, R. (2006), *Fault-diagnosis systems: An introduction from fault detection to fault tolerance.*

Jaffe, A. B. & Stavins, R. N. (1994), 'The energy-efficiency gap What does it mean?', *Energy Policy* **22**(10), 804–810.

Jain, R. K., Smith, K. M., Culligan, P. J. & Taylor, J. E. (2014), 'Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy', *Applied Energy* **123**, 168–178.

James, G., Witten, D., Tibshirani, R. & Hastie, T. (2013), *An Introduction to Statistical Learning with Applications in R.*

Jayaweera, T., Stern, F., Violette, D. M. & Baumgartner, R. (2013), The Uniform Methods Project : Methods for Determining Energy Efficiency Savings for Specific Measures The Uniform Methods Project, Technical Report April, National Renewable Energy Laboratory, US DOE.

Jovanović, B. & Filipović, J. (2016), 'ISO 50001 standard-based energy management maturity model - Proposal and validation in industry', *Journal of Cleaner Production* **112**, 2744–2755.

Karatasou, S., Santamouris, M. & Geros, V. (2006), 'Modeling and predicting building's energy use with artificial neural networks: Methods and results', *Energy and Buildings* **38**(8), 949–958.

Katipamula, S. & Brambley, M. R. (2005*a*), 'Review article: Methods for fault detection, diagnostics, and prognostics for building systems—a review, part II', *HVAC and R Research* **11**(2), 169–187.

Katipamula, S. & Brambley, M. R. (2005*b*), 'Review article: Methods for fault detection, diagnostics, and prognostics for building systems—a review, part II', *HVAC and R Research* **11**(2), 169–187.

Ke, M. T., Yeh, C. H. & Jian, J. T. (2013), 'Analysis of building energy consumption parameters and energy savings measurement and verification by applying eQUEST software', *Energy and Buildings* **61**, 100–107.

Ke, M.-T., Yeh, C.-H. & Su, C.-J. (2017), 'Cloud computing platform for real-time measurement and verification of energy performance', *Applied Energy* **188**, 497–507.

Kelly Kissock, J. & Eger, C. (2008), 'Measuring industrial energy savings', *Applied Energy* **85**(5), 347–361.

Klemick, H. & Wolverton, A. (2013), Energy-efficiency gap, *in* J. F. Shogren, ed., 'Encyclopedia of energy, natural resource and environmental economics', Vol. 1, Elsevier, pp. 74–81.

Kohavi, R. (1995), A study of cross-validation and bootstrap for accuracy estimation and model selection, *in* 'International Joint Conference on Artificial Intelligence', Vol. 14, pp. 1137–1143.

Kreider, J. F., Claridge, D. E., Curtiss, P., Dodier, R., Haberl, J. S. & Krarti, M. (1995), 'Building Energy Use Prediction and System Identification Using Recurrent Neural Networks', *Journal of Solar Energy Engineering* **117**(3), 161.

Kreider, J. & Haberl, J. (1994*a*), 'Predicting hourly building energy usage: The results of the 1993 ASHRAE Great Energy Predictor Shootout identify the most accurate method for making hourly energy use predictions', *ASHRAE Journal* pp. 72–81.

Kreider, J. & Haberl, J. (1994*b*), Predicting hourly building energy use: The great energy predictor shootout – Overview and discussion of results, *in* '1994 American Society of Heating, Refrigerating, and Air Conditioning Engineers (ASHRAE) annual meeting', American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., Atlanta, GA (United States).

Kumari, B. & Swarnkar, T. (2011), 'Filter versus Wrapper Feature Subset Selection in Large Dimensionality Microarray : A Review', *International Journal of Computer Science and Information Technologies* **2**(3), 1048–1053.

Kupser, J., Francois, S., Rego, J., Steele-Mosey, P., Galvin, T. & McDonald, C. (2016), M&V 2.0: Hype vs. Reality, *in* 'ACEEE Summer Study on Energy Efficiency in Buildings 2016', Pacific Groive, California.

Langlois-Bertrand, S., Benhaddadi, M., Jegen, M. & Pineau, P.-O. (2015), 'Political-institutional barriers to energy efficiency', *Energy Strategy Reviews* **8**, 30–38.

Lee, P., Lam, P. T. I. & Lee, W. L. (2015), 'Risks in Energy Performance Contracting (EPC) projects', *Energy and Buildings* **92**, 116–127.

Lee, S., Painter, F. & Claridge, D. E. (2007), 'Whole-Building Commercial HVAC System Simulation for Use in Energy Consumption Fault Detection', *ASHRAE Transactions* **113**(2), 52–61.

Li, Q., Meng, Q., Cai, J., Yoshino, H. & Mochida, A. (2009*a*), 'Applying support vector machine to predict hourly cooling load in the building', *Applied Energy* **86**(10), 2249–2256.

Li, Q., Meng, Q., Cai, J., Yoshino, H. & Mochida, A. (2009*b*), 'Predicting hourly cooling load in the building: A comparison of support vector machine and different artificial neural networks', *Energy Conversion and Management* **50**(1), 90–96.

Li, Q., Ren, P. & Meng, Q. (2010), Prediction model of annual energy consumption of residential buildings, *in* 'International Conference on Advances in Energy Engineering', pp. 223–226.

Liang, X., Hong, T. & Shen, G. Q. (2016), 'Improving the accuracy of energy baseline models for commercial buildings with occupancy data', *Applied Energy* **179**, 247–260.

Ljung, L. (1987), *System identification: theory for the user*, Prentice-hall.

Lü, X., Lu, T., Kibert, C. J. & Viljanen, M. (2014), 'A novel dynamic modeling approach for predicting building energy performance', *Applied Energy* **114**, 91–103.

Malinowski, A., Cholewo, T. & Zurada, J. (1995), Capabilities and limitations of feedforward neural networks with multilevel neurons, *in* 'Proceedings of ISCAS'95 - International Symposium on Circuits and Systems', Vol. 1, pp. 131–134.

Maritz, J., Lubbe, F., Lagrange, L., Maritz, J., Lubbe, F. & Lagrange, L. (2018), 'A Practical Guide to Gaussian Process Regression for Energy Measurement and Verification within the Bayesian Framework', *Energies* **11**(4), 935.

McKane, A., Desai, D., Matteini, M., Williams, R. & Risser, R. (2009), 'Thinking Globally : How ISO 50001 – Energy Management can make industrial energy efficiency standard practice', *Environmental energy technologies* pp. 1–16.

McKane, A., Therkelsen, P., Scodel, A., Rao, P., Aghajanzadeh, A., Hirzel, S., Zhang, R., Prem, R., Fossa, A., Lazarevska, A. M., Matteini, M., Schreck, B., Allard, F., Villegal Alcántar, N., Steyn, K., Hürdoğan, E., Björkman, T. & O'Sullivan, J. (2017), 'Predicting the quantifiable impacts of ISO 50001 on climate change mitigation', *Energy Policy* **107**, 278–288.

Meath, C., Linnenluecke, M. & Griffiths, A. (2016), 'Barriers and motivators to the adoption of energy savings measures for small- and medium-sized enterprises (SMEs): the case of the ClimateSmart Business Cluster program', *Journal of Cleaner Production* **112**, 3597–3604.

Neto, A. H. & Fiorelli, F. A. S. (2008), 'Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption', *Energy and Buildings* **40**(12), 2169–2176.

Ngai, E. W., Chau, D. C., Poon, J. K. & To, C. K. (2013), 'Energy and utility management maturity model for sustainable manufacturing process', *International Journal of Production Economics* **146**(2), 453–464.

Nielsen, H. A. & Madsen, H. (2006), 'Modelling the heat consumption in district heating systems using a grey-box approach', *Energy and Buildings* **38**(1), 63–71.

O'Donovan, P., Gallagher, C., Bruton, K. & O'Sullivan, D. T. (2018), 'A fog computing industrial cyber-physical system for embedded low-latency machine learning Industry 4.0 applications', *Manufacturing Letters* .

O'Donovan, P., Leahy, K., Bruton, K. & O'Sullivan, D. T. J. (2015), 'An industrial big data pipeline for data-driven analytics maintenance applications in large-scale smart manufacturing facilities', *Journal of Big Data* **2**(1), 25.

Olinga, Z., Xia, X. & Ye, X. (2017), 'A cost-effective approach to handle measurement and verification uncertainties of energy savings', *Energy* **141**, 1600–1609.

Paris, B., Eynard, J., Grieu, S. & Polit, M. (2011), 'Hybrid PID-fuzzy control scheme for managing energy resources in buildings', *Applied Soft Computing* **11**(8), 5068–5080.

Paris, B., Eynard, J., Grieu, S., Talbert, T. & Polit, M. (2010), 'Heating control schemes for energy management in buildings', *Energy and Buildings* **42**(10), 1908–1917.

Perroni, M. G., Gouvea da Costa, S. E., Pinheiro de Lima, E., Vieira da Silva, W. & Tortato, U. (2018), 'Measuring energy performance: A process based approach', *Applied Energy* **222**, 540–553.

Prashar, A. (2017), 'Energy efficiency maturity (EEM) assessment framework for energy-intensive SMEs: Proposal and evaluation', *Journal of Cleaner Production* **166**, 1187–1201.

Project Haystack (2018).
URL: *http://project-haystack.org/about*

Rabl, A. & Rialhe, A. (1992), 'Energy signature models for commercial buildings: test with measured data and interpretation', *Energy and Buildings* **19**(2), 143–154.

Rahman, A., Srikumar, V. & Smith, A. D. (2018), 'Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks', *Applied Energy* **212**, 372–385.

Reddy, T. A. (2006), 'Literature review on calibration of building energy simulation programs', *ASHRAE Transactions* **112**(1), 226–240.

Robinson, C., Dilkina, B., Hubbs, J., Zhang, W., Guhathakurta, S., Brown, M. A. & Pendyala, R. M. (2017), 'Machine learning approaches for estimating commercial building energy consumption', *Applied Energy* **208**, 889–904.

Rohdin, P. & Thollander, P. (2006), 'Barriers to and driving forces for energy efficiency in the non-energy intensive manufacturing industry in Sweden', *Energy* **31**(12), 1836–1844.

Rohdin, P., Thollander, P. & Solding, P. (2007), 'Barriers to and drivers for energy efficiency in the Swedish foundry industry', *Energy Policy* **35**(1), 672–677.

Rossi, F. & Velázquez, D. (2015), 'A methodology for energy savings verification in industry with application for a CHP (combined heat and power) plant', *Energy* **89**, 528–544.

Rossi, F., Velázquez, D., Monedero, I. & Biscarri, F. (2014), 'Artificial neural networks and physical modeling for determination of baseline consumption of CHP plants', *Expert Systems with Applications* **41**(10), 4658–4669.

Royer, S., Bressan, M., Thil, S. & Talbert, T. (2013), 'Modelling of a multi-zone building and assessment of its thermal behaviour using an energy simulation software', *2013 IEEE International Conference on Automation Science and Engineering (CASE)* pp. 735–740.

Russell, E. L., Chiang, L. H. & Braatz, R. D. (2012), *Data-driven Methods for Fault Detection and Diagnosis in Chemical Processes*.

Saeys, Y., Inza, I. & Larrañaga, P. (2007), 'A review of feature selection techniques in bioinformatics', *Bioinformatics* **23**(19), 2507–2517.

Saidur, R., Rahim, N. A. & Hasanuzzaman, M. (2010), 'A review on compressed-air energy use and energy savings', *Renewable and Sustainable Energy Reviews* **14**(4), 1135–1153.

Seth, Y. (2017), 'Bootstrapping – A Powerful Resampling Method in Statistics – Let the Machines Learn'.

URL: *https://yashuseth.blog/2017/12/02/bootstrapping-a-resampling-method-in-statistics/*

Sever, F., Kissock, K., Brown, D. & Mulqueen, S. (2011), Estimating industrial building energy savings using inverse simulation, *in* 'ASHRAE Transactions', Vol. 117, pp. 348–355.

Sjöberg, J., Zhang, Q., Ljung, L., Benveniste, A., Delyon, B., Glorennec, P. Y., Hjalmarsson, H. & Juditsky, A. (1995), 'Nonlinear black-box modeling in system identification: a unified overview', *Automatica* **31**(12), 1691–1724.

Skumatz, L. A., Khawaja, M. & Colby, J. (2009), 'Lessons Learned and Next Steps in Energy Efficiency Measurement and Attribution: Energy Savings, Net to Gross, Non-Energy Benefits, and Persistence of Energy Efficiency Behavior', (November 2009), 166.

Smarra, F., Jain, A., de Rubeis, T., Ambrosini, D., D'Innocenzo, A. & Mangharam, R. (2018), 'Data-driven model predictive control using random forests for building energy optimization and climate control', *Applied Energy* **226**, 1252–1272.

Smola, A and Schölkopf, B. (2004), 'Tutorial on Support Vector Regression', *Statistics and Computing* **14**(3), 199–222.

Song, Y.-h., Akashi, Y. & Yee, J. J. (2008), 'A development of easy-to-use tool for fault detection and diagnosis in building air-conditioning systems', *Energy and Buildings* **40**(2), 71–82.

Sorrell, S. (2007), *The Rebound Effect: an assessment of the evidence for economy-wide energy savings from improved energy efficiency*, Vol. 42.

Sorrell, S. (2009), 'Jevons' Paradox revisited: The evidence for backfire from improved energy efficiency', *Energy Policy* **37**(4), 1456–1469.

Sorrell, S., Schleich, J., Scott, S., Eoin, O., Fergal, T., Ulla, B., Katrin, O. & Peter, R. (2000), Reducing barriers to energy efficiency in public and private organisations, Technical report.

Stergiou, C. & Siganos, D. (2013), 'Neural Networks'.
URL: *www.doc.ic.ac.uk/ nd/surprise_96/journal/vol4/cs11/report.html*

Sustainable Energy Authority of Ireland (2017), 'Energy in Ireland 1990-2016: 2017 Report'.

Tang, F., Kusiak, A. & Wei, X. (2014), 'Modeling and short-term prediction of HVAC system with a clustering algorithm', *Energy and Buildings* **82**, 310–321.

Tardioli, G., Kerrigan, R., Oates, M., Donnell, J. O. & Finn, D. (2015), Data driven approaches for prediction of building energy consumption at urban level, *in* 'Energy Procedia', Vol. 78, pp. 3378–3383.

Teeter, J. & Chow, M. Y. (1998), 'Application of functional link neural network to hvac thermal dynamic system identification', *IEEE Transactions on Industrial Electronics* **45**(1), 170–176.

Therkelsen, P., McKane, A., Sabouni, R., Evans, T. & Scheihing, P. (2013), 'Assessing the Costs and Benefits of the Superior Energy Performance Program', *2013 ACEEE Summer Study on Energy Efficiency in Industry* (July), 14.

Therkelsen, P., Rao, P., Sholes, D., Meffert, B., Green, R., Nimbalkar, S. U. & Mckane, A. (2016), The Value of Regression Models in Determining Industrial Energy Savings, *in* 'ECEEE Industrial Efficiency Conference', Berlin, Germany.

Therkelsen, P., Scheihing, P. & Siciliano, G. (2018), Establishing a platform to harmonize ISO 50001 energy performance improvement measurement and verification protocols, *in* 'Proceedings of ECEEE Industrial Efficiency 2018', European Commission for an Energy Efficient Economy, Berlin, Germany.

Thollander, P., Backlund, S., Trianni, A. & Cagno, E. (2013), 'Beyond barriers - A case study on driving forces for improved energy efficiency in the foundry industries in Finland, France, Germany, Italy, Poland, Spain, and Sweden', *Applied Energy* **111**, 636–643.

Thollander, P. & Ottosson, M. (2008), 'An energy efficient Swedish pulp and paper industry - Exploring barriers to and driving forces for cost-effective energy efficiency investments', *Energy Efficiency* **1**(1), 21–34.

Touzani, S., Granderson, J. & Fernandes, S. (2018), 'Gradient boosting machine for modeling the energy consumption of commercial buildings', *Energy and Buildings* **158**, 1533–1543.

Trianni, A., Cagno, E. & Farné, S. (2016), 'Barriers, drivers and decision-making process for industrial energy efficiency: A broad study among manufacturing small and medium-sized enterprises', *Applied Energy* **162**, 1537–1551.

United Nations Framework Convention on Climate Change (2015), 'Paris Climate Change Conference-November 2015, COP 21'.

US Department of Energy (2012), 'Superior Energy Performance Measurement and Verification Protocol for Industry', *SEP resource* .

US Department of Energy (2017), 'BEST Directory | Building Energy Software Tools'.

US Department of Energy (2018), 'Superior Energy Performance | Department of Energy'.

Vapnik, V. (2000), *The nature of statistical learning - 2nd edition*, 2nd editio edn, Springer-Verlag, New York.

Walter, T., Price, P. N. & Sohn, M. D. (2014), 'Uncertainty estimation improves energy measurement and verification procedures', *Applied Energy* **130**, 230–236.

Wang, J., Klinger, K., Appicello, A., Ash, J., Brehem, M., Coulson, C., Duke, E., Henderson, J., Hrobak, C., Neidig, J., Snavely, J. & Woznicki, S. (2009), *Encyclopedia of Data Warehousing and Mining, Second Edition*.

Wang, L. & Wong, N. H. (2008), 'Coupled simulations for naturally ventilated residential buildings', *Automation in Construction* **17**(4), 386–398.

Wang, S. & Xu, X. (2006), 'Simplified building model for transient thermal performance estimation using GA-based parameter identification', *International Journal of Thermal Sciences* **45**(4), 419–432.

Wang, X., Chen, Z., Yang, C. & Chen, Y. (1999), 'Gray predicting theory and application of energy consumption of building heat-moisture system', *Building and Environment* **34**(4), 417–420.

Wang, Z. & Srinivasan, R. S. (2017), 'A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models', *Renewable and Sustainable Energy Reviews* **75**, 796–808.

Wei, X., Kusiak, A., Li, M., Tang, F. & Zeng, Y. (2015), 'Multi-objective optimization of the HVAC (heating, ventilation, and air conditioning) system performance', *Energy* **83**, 294–306.

Whiten, B. (2013), 'Model completion and validation using inversion of grey box models', *ANZIAM Journal* **54**, 187–199.

Wong, S., Wan, K. K. & Lam, T. N. (2010), 'Artificial neural networks for energy analysis of office buildings with daylighting', *Applied Energy* **87**(2), 551–557.

Worden, K., Wong, C., Parlitz, U., Hornstein, A., Engster, D., Tjahjowidodo, T., Al-Bender, F., Rizos, D. & Fassois, S. (2007), 'Identification of pre-sliding and sliding friction dynamics: Grey box and black-box models', *Mechanical Systems and Signal Processing* **21**(1), 514–534.

Wu, D., Liu, S., Zhang, L., Terpenny, J., Gao, R. X., Kurfess, T. & Guzzo, J. A. (2017), 'A fog computing-based framework for process monitoring and prognosis in cyber-manufacturing', *Journal of Manufacturing Systems* **43**, 25–34.

Wulandari, M., Laskurain, I., Fa, M. C. & Heras-Saizarbitoria, I. (2015), *Early Adoption of ISO 50001 Standard: An Empirical Study*, Springer International Publishing, Cham, pp. 183–202.

Xia, X. & Zhang, J. (2013), 'Mathematical description for the measurement and verification of energy efficiency improvement', *Applied Energy* **111**, 247–256.

Xu, X., Taylor, J. E., Pisello, A. L. & Culligan, P. J. (2012), 'The impact of place-based affiliation networks on energy conservation: An holistic model that integrates the influence of buildings, residents and the neighborhood context', *Energy and Buildings* **55**, 637–646.

Yao, R. & Steemers, K. (2005), 'A method of formulating energy load profile for domestic buildings in the UK', *Energy and Buildings* **37**(6), 663–671.

Ye, X. & Xia, X. (2016), 'Optimal metering plan for measurement and verification on a lighting case study', *Energy* **95**, 580–592.

Yik, F. W., Burnett, J. & Prescott, I. (2001), 'Predicting air-conditioning energy consumption of a group of buildings using different heat rejection methods', *Energy and Buildings* **33**(2), 151–166.

Yildiz, B., Bilbao, J. I. & Sproul, A. B. (2017), 'A review and analysis of regression and machine learning models on commercial building electricity load forecasting', *Renewable and Sustainable Energy Reviews* **73**, 1104–1122.

Yu, Z., Haghighat, F., Fung, B. C. & Yoshino, H. (2010), 'A decision tree method for building energy demand modeling', *Energy and Buildings* **42**(10), 1637–1646.

Zhang, Y., O'Neill, Z., Dong, B. & Augenbroe, G. (2015), 'Comparisons of inverse modeling approaches for predicting building energy performance', *Building and Environment* **86**, 177–190.

Zhang, Y. & Yang, Y. (2015), 'Cross-validation for selecting a model selection procedure', *Journal of Econometrics* **187**(1), 95–112.

Zhao, H. X. & Magoulès, F. (2012), 'A review on the prediction of building energy consumption', *Renewable and Sustainable Energy Reviews* **16**(6), 3586–3592.

Zhao, Y., Wen, J. & Wang, S. (2015), 'Diagnostic Bayesian networks for diagnosing air handling units faults - Part II: Faults in coils and sensors', *Applied Thermal Engineering* **90**, 145–157.

Zhao, Y., Wen, J., Xiao, F., Yang, X. & Wang, S. (2017), 'Diagnostic Bayesian networks for diagnosing air handling units faults – part I: Faults in dampers, fans, filters and sensors', *Applied Thermal Engineering* **111**, 1272–1286.

Zhou, Q., Wang, S., Xu, X. & Xiao, F. (2008), 'A grey-box model of next-day building thermal load prediction for energy-efficient control', *International Journal of Energy Research* **32**(15), 1418–1431.

Appendix A

Intelligent M&V: Graphical User Interface

The following images are samples of the graphical user interface of the IntelliMaV application.

Project Parameters

Data Input

Project Description | Input Data | Feature Selection | Data Cleaning | Model Construction

Step 2: Connect to Facility Data Source

Retrieve data from cloud-based storage facility

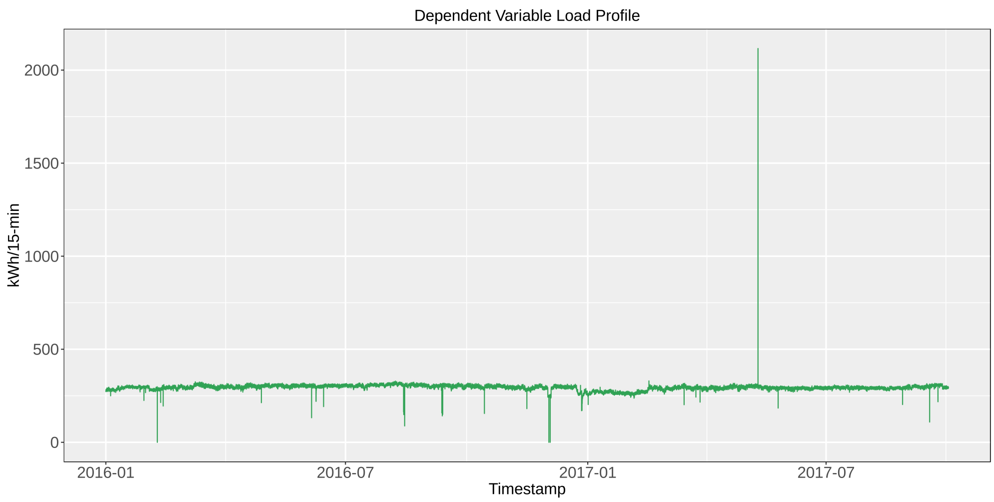
Input File Source https://s3-eu-west-1.amazona

Gather Data & Show Summary Stats

Preview of data

Timestamp	Line.43.Compressed.Air	Line.44.Compressed.Air	Line.45.Compressed.Air	F1.N4311.Low.Purity	F1.N4511.low.purity	F1.N4411.low.puri
2016-10-05 00:15:00	115	115	94	150	164	134
2016-10-05 00:30:00	120	114	94	150	163	135
2016-10-05 00:45:00	115	112	82	150	163	134
2016-10-05 01:00:00	116	118	83	150	163	135
2016-10-05 01:15:00	114	113	92	150	163	134
2016-10-05 01:30:00	113	112	93	150	163	134

Load profile



Continue

Feature Selection

Project Description   Input Data   Feature Selection   Data Cleaning   Model Construction

Step 3: Identify Relevant Independent Variables

The data set needs to be refined in order to optimise the model construction process. It is important that only those variables that offer importance in model construction are brought forward for analysis, i.e statistically significant variables must be identified and designated *"features"* status. A novel feature selection algorithm has been designed to achived this. The algorithm takes the following form:

1. The Spearman rank correlation (***p***) is calculated for each input variable.
2. Variables are ordered by decreasing ***p***.
3. A subset is created with the dependent variable and variable with the highest ***p***.
4. An ordinary least square regression model is trained using the subset.
5. The initial r^2 is calculated.
6. If the initial r^2 is greater than 0.01 then the initial subset is brought forward for analysis. If not, the variable is disregarded.
7. Iteratively, variables are added to the subset. Those that increase the initial r^2 by greater than 0.01 stay in the subset. Those that do not are disregarded.
8. A final subset, known as the **feature set** is identified.

Apply Feature Selection Algorithm

Feature Set & Relevant Statistical Metrics

Feature	Mean	No. Unique	No. Missing	Min	1st Quartile	Median	3rd Quartile	Max
MDB.A.Cub.A2.Bus.Coupler	77.2319	264	12	0	0	0	203	359
MDB.C.Cub.M2.MCC.5..WP.part.of.Phase.1.	20.4738	18	12	0	20	20	21	148
MDB.C.Cub.K2.SPARE	0.0023	6	12	0	0	0	0	54
MDB.K.Cub.K25.HPN.2.Feeder	92.822	28	19	0	93	93	94	654
MDB.E.Cub.E09.MLP.3	7.0501	29	35	0	7	7	8	54
FI.N0208.HPN.2.Phase.5	253.7574	245	44	0	256	258	260	1820
hvac (Dependent Variable)	294.8121	162	0	0	290	296	303	2116

Data Cleaning

Project Description   Input Data   Feature Selection   Data Cleaning   Model Construction

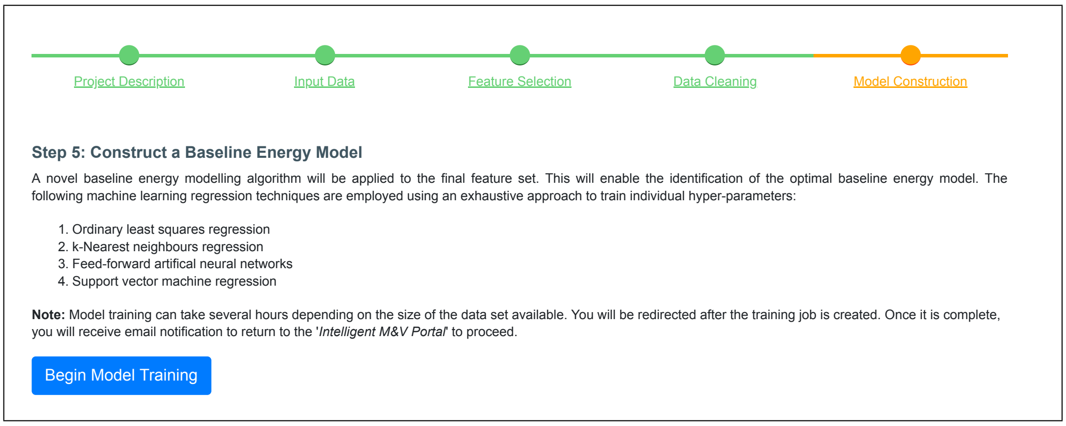
Step 4: Data Availability Assessment & Cleaning

Data cleaning is the process of detecting and removing inaccurate entries in a data set. It is required to ensure quality and integrity are present in the data set used to construct the baseline energy models. It can also increase the applicability of the model in the reporting period. To ensure transparency in the process, the following approach is taken to clean the raw data:

1. As per the IPMVP, baseline data should not be replaced by modelled data, except when using Option D (EVO, IPMVP Vol. I, 2012). Therefore, **no backfilling is carried out**.
2. The scope of data cleaning in this application is limited to simply identifying unclean data and subsequently removing it.
3. Variables that are not of sufficient quality are identifed using the statistical measures in previous step. Features with more than **5%** of poor quality data should be omitted entirely from the subset.
4. Any features that fall short of this 5% omission threshold are cleaned by omitting periods of unclean data.
5. The final feature set is stored in an open-source relational database for documentation purposes.

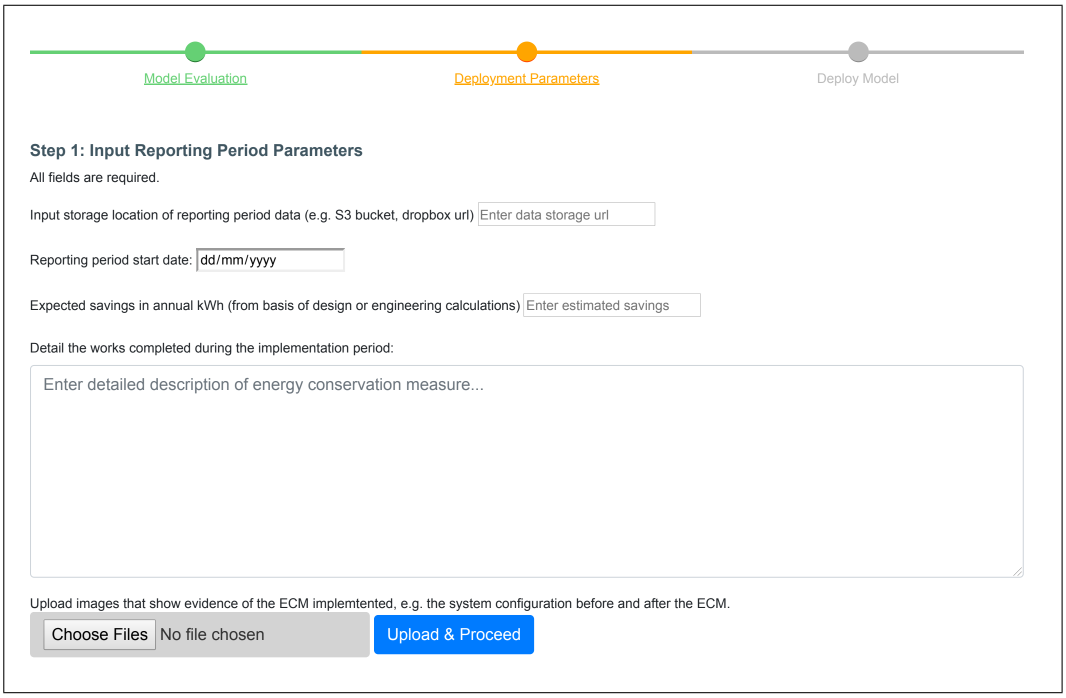
Apply Data Cleaning Algorithm

Model Construction



Model Evaluation

Deployment Parameters



Final Deployment